(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 October 2001 (18.10.2001)

PCT

(10) International Publication Number
WO 01/77350 A2

(51) International Patent Classification[7]: C12N 15/79

(21) International Application Number: PCT/US01/11436

(22) International Filing Date: 4 April 2001 (04.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/545,574 7 April 2000 (07.04.2000) US

(71) Applicant *(for all designated States except US)*: **LARGE SCALE BIOLOGY CORPORATION** [US/US]; 3333 Vaca Valley Parkway, Vacaville, CA 95688 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **PALMER, Kenneth, E.** [ZA/US]; 707 West Monte Vista Avenue, Vacaville, CA 95688 (US). **POGUE, Gregory, P.** [US/US]; 419 Trillick Court, Vacaville, CA 95688 (US).

(74) Agent: **HALLUIN, Albert, P.**; Howrey Simon Arnold & White, 301 Ravenswood Avenue, Box 34, Menlo Park, CA 94025 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/77350 A2

(54) Title: COMPOSITIONS AND METHODS FOR INHIBITING GENE EXPRESSION

(57) Abstract: The present invention provides an eukaryotic recombinant vector suited for bi-directional transcription of a transgene to yield both sense and antisense RNA transcripts of the transgene in an eukaryotic cell. The invention vectors are particularly suited for mediating gene silencing in a variety of biological systems. The present invention also provides host cells and transgenic plants comprising the invention vectors. Further provided by the invention are methods of inhibiting expression of an endogenous gene present in an eukaryotic cell. Also included is a method of identifying a biological function(s) of an endogenous gene of interest in an eukaryotic cell by selectively inhibiting the expression of the endogenous gene.

COMPOSITIONS AND METHODS FOR INHIBITING GENE EXPRESSION

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of U.S. Patent Application 09/545,574, filed April 7, 2000, pending, which is hereby incorporated herein by reference in its entirety.

STATEMENT OF RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH

Not applicable.

TECHNICAL FIELD

This invention is in the field of genetic analysis. Specifically, the invention relates to the generation of a eukaryotic vector that allows bi-directional transcription of a transgene to yield both sense and antisense RNA transcripts from the same transgene. The compositions and methods embodied in the present invention are particularly useful for targeted inhibition of gene expression in a eukaryotic cell.

BACKGROUND OF THE INVENTION

The structure and biological behavior of a cell is determined by the pattern of gene expression within that cell at a given time. Perturbations of gene expression have long been acknowledged to account for a vast number of diseases including, numerous forms of cancer, vascular diseases, neuronal and endocrine diseases. Abnormal expression patterns, in form of amplification, deletion, gene rearrangements, and loss or gain of function mutations, are now known to lead to aberrant behavior of a disease cell. Aberrant gene expression has also been noted as a defense mechanism of certain organisms to ward off the threat of pathogens.

One of the major challenges of genetic engineering has been to regulate the expression of targeted genes that are implicated in a wide diversity of physiological responses. While overexpression of an exogenously introduced transgene in a eukaryotic cell is relatively straightforward, targeted inhibition of specific genes has been more difficult to achieve. Traditional approaches for suppressing gene expression, including site-directed gene disruption, antisense RNA or co-suppressor injection, require complex genetic manipulations or heavy dosages of suppressors that often exceeds the toxicity tolerance level of the host cell.

Recently, a new technique, "double-stranded RNA interference" has emerged in the study of gene silencing. Several research groups have demonstrated a marked inhibition of a specific nuclear gene expression in a wide range of eukaryotes by introduction into cells of dsRNA fragments that bear sequence homology with the nuclear gene. For instance, Fire et al. (1998) *Nature* **395**: 854 reported the success of gene-specific interference in *C. elegans* that was mediated by ingested *E. coli* carrying a prokaryotic vector capable of producing both sense and antisense RNAs of the selected *C. elgans* genes. Misquitta et al. demonstrated the targeted disruption of *nautilus* gene in *Drosophila melanogaster* by injecting into the Drosophila embryo multiple copies of *nautilus* dsRNA. See Misquitta et al. (1999) *PNAS U.S.A.* 96:1451-1456. Studies by Ngô et al. (1998) *Proc. Natl. Acad. of Sci. U.S.A.*, **96**:1451-1456 confirmed that dsRNA interference also occurs in certain protozoan species. Earlier studies by Cogoni et al. and Hamilton et al. suggested that formation of dsRNA play a pivotal role in gene silencing in fungi *Neurospora crassa* and other plants. See Cogoni et al. (1999) *Nature* **399**: 166-169; Hamilton et al. (1999) *Science* **286**: 950-952; and Waterhouse et al. (1999) *PNAS U.S.A.* **95**: 13959-13964. More recent investigations by Wargelius et al. revealed that this phenomenon is also conserved in vertebrates such as the zebrafish. Wargelius et al. *Biochem. Biophys. Res. Commun.* **263**: 156-161.

Current techniques for achieving RNA mediated gene silencing include: (a) use of prokaryotic vectors capable of transcribing both sense and antisense RNA (Fire et al. (1998) *Nature* **395**: 854; (b) *in vitro* transcription of individual strands of a selected gene followed by annealing the transcribed sense and antisense RNAs (see, e.g. Misquitta et al. (1999) *PNAS U.S.A.* 96:1451-1456); and possibly (c) viruses induced gene silencing (see, e.g. Angell et al. (1997) *EMBO Journal* **16**:

3675-3684; Angell et al. (1999) *Plant Journal* **20**: 357-362). However, these methods bear a number of intrinsic limitations. First, none of these methods employs gene delivery vehicles that are applicable for consistent and persistent inhibition of gene expression in a eukaryote. Second, these existing methods do not necessarily result in production of a substantially homogenous population of dsRNAs. Notably, the *in vitro* preparation of double-stranded RNAs by transcribing and annealing sense RNA transcripts to antisense transcripts is time consuming, labor intensive, and not amenable for mass production or high-throughput analyses.

Thus, there remains a considerable need for compositions and methods to effect dsRNA-mediated gene silencing. An ideal reagent would be a self-replicating vector that is (a) capable of autonomous replication and expression of a selected transgene in a eukaryotic cell; and (b) capable of yielding both sense and antisense RNA transcripts from the same transgene, so as to effect production of dsRNA transcripts in a eukaryotic host cell. The present invention satisfies these needs and provides related advantages as well.

SUMMARY OF THE INVENTION

A principal aspect of the present invention is the design of a eukaryotic recombinant vector to effect gene silencing in a eukaryotic cell that is susceptible to dsRNA-mediated reduction of gene expression. Such a vector allows bi-directional transcription of a transgene to yield both sense and antisense RNA transcripts of the same transgene in a eukaryotic cell. While not being bound to any one theory, the production of dsRNAs induces transcriptional and/or post-transcriptional gene silencing in the host cell. Accordingly, the present invention provides a recombinant vector having the following unique characteristics: it comprises a viral replicon having two overlapping transcription units arranged in an opposing orientation and flanking a transgene of interest, wherein the two overlapping transcription units yield both sense and antisense RNA transcripts from the same transgene fragment in a eukaryotic host cell.

In one aspect of this embodiment, each of the overlapping transcription units of the vector comprises a promoter and a terminator that are arranged in one of the configurations shown in Figure 2(a)-(d). The promoter can be constitutive or

inducible; it can be active in all tissues and cell types of an organism or operative only in selected tissues (i.e. tissue-specific).

In another aspect, the recombinant vector comprises a viral replicon that is derived from a DNA virus. Such DNA viruses can be selected from the group consisting of *Geminivirus, Caulimoviridae, Badnaviridae, Circoviridae, Circinoviridae, Parvoviridae, Papovaviridae, Polyomaviridae, Adenoviridae, Herpesviridae, Poxviridae, Iridoviridae, Baculoviridae, Hepadnaviridae, Retroviridae, Gyrovirus, Nanovirus*, and African Swine Fever virus.

In yet another aspect, the subject vector is capable of autonomous replication in a eukaryotic cell.

In still another aspect, the subject vector is capable of inhibiting expression of genes endogenous to a eukaryotic host cell. Non-limiting representative eukaryotic cells whose gene expression can be inhibited upon introduction of the subject vectors are fungi, yeast cells, plant cells, inset, avian, mammalian or other animal cells. Preferably, the vectors effect a reduced expression of an endogenous gene that is substantially homologous to the transgene contained in the overlapping transcription units of the vectors. More preferably, delivery of the vectors into a suitable host cell results in a phenotypic change of the host cell. In certain preferred embodiments, the endogenous gene is native to the host cell. The endogenous gene can also be heterologous to the host cell. In some embodiments, the endogenous gene is a pathogenic gene derived from one or more members of the group consisting of virus, bacterium, fungus, and protozoa. The transgene carried in the vector can be a nucleotide sequence that encodes a membrane protein, a cytosolic protein, a secreted protein, a nuclear protein, or a chaperon protein.

The present invention also provides host cells transformed with the invention vectors. The present invention further provides a transgenic plant comprising a eukaryotic recombinant vector of the present invention.

Also provided by the present invention is a kit for generating a double-stranded RNA transcript in a eukaryotic cell that contains the subject vectors in suitable packaging.

Further embodied in the present invention is a method of inhibiting expression of an endogenous gene present in a eukaryotic cell. The method involves: (a) providing a eukaryotic recombinant vector containing a transgene

that is substantially homologous to the endogenous gene; (b) introducing the eukaryotic recombinant vector into the eukaryotic cell; and (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene that is contained in the transcription units of the vector, and thereby inhibiting expression of the corresponding endogenous gene in the eukaryotic cell.

Also included in the present invention is a method of identifying a biological function(s) of an endogenous gene of interest in a eukaryotic cell by selectively inhibiting the expression of the endogenous gene. The method comprises: (a) providing a eukaryotic recombinant vector containing a transgene that is substantially homologous to the endogenous gene; (b) introducing the eukaryotic recombinant vector of (a) into the eukaryotic cell; (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene contained in the eukaryotic recombinant vector and thereby inhibiting expression of the endogenous gene in the eukaryotic cell; and (d) determining one or more phenotypic changes in the eukaryotic cell that correlate with the inhibited expression of the endogenous gene, thereby identifying the biological function(s) of the endogenous gene in the eukaryotic cell. In essence, the subject methods allow the creation of a transient or more long-term gene-specific knock-out system for analyzing the biological function of any endogenous gene of interest.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic representation of the process for production of dsRNA transcripts by a subject vector containing two overlapping transcription units.

Figure 2 (a)-(d) depict four different configurations of the overlapping transcription units of the subject vectors.

Figure 3 is a schematic representation of an exemplary construct MSVLSB-6.

Figure 4 depicts the nucleotide sequence of the vector pMSVLSB-1 (SEQ ID NO:9) described in Examples 1-2.

Figure 5 depicts the nucleotide sequence of the vector pMSVLSB-2 (SEQ ID NO:10) described in Examples 1-2.

Figure 6 depicts the nucleotide sequence of the vector pMSVLSB-3 (SEQ ID NO:11) described in Examples 1-2.

Figure 7 depicts the nucleotide sequence of the vector pMSVLSB-4 (SEQ ID NO:12) described in Examples 1-2.

Figure 8 depicts the nucleotide sequence of the vector pMSVLSB-5 (SEQ ID NO:13) described in Examples 1-2.

Figure 9 depicts the nucleotide sequence of the vector pMSVLSB-6 (SEQ ID NO: 14) described in Examples 1-2.

MODES FOR CARRYING OUT THE INVENTION

Throughout this disclosure, various publications, patents and published patent specifications are referenced by an identifying citation. The disclosures of these publications, patents and published patent specifications are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

General Techniques:

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See, *e.g.,* Matthews, PLANT VIROLOGY, 3rd edition (1991); Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R.I. Freshney, ed. (1987)).

As used in the specification and claims, the singular form "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a cell" includes a plurality of cells, including mixtures thereof.

Definitions:

A "plant cell" refers to the structural and physiological unit of plants, consisting of a protoplast and the cell wall.

A "protoplast" is an isolated cell without cell walls, having the potency for regeneration into cell culture, tissue or whole plant.

A "host cell" includes an individual cell or cell culture which can be or has been a recipient for vector(s) or for incorporation of nucleic acid molecules and/or proteins. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in genomic of total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation. A host cell includes cells transfected *in vivo* with a polynucleotide(s) of this invention.

The terms "polynucleotide", "nucleotides" and "oligonucleotides" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

A "gene" refers to a polynucleotide containing at least one open reading frame that is capable of encoding a particular protein after being transcribed and translated.

"Genes of a specific developmental origin" refer to genes expressed at certain but not all developmental stages. For instance, a gene may be of embryonic or adult origin depending on the stage during which the gene is expressed.

A "disease-associated" or "disease-causing" gene refers to any gene which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at normal or abnormal level.

A gene "database" denotes a set of stored data which represent a collection of sequences including nucleotide and peptide sequences, which in turn represent a collection of biological reference materials.

As used herein, "expression" refers to the process by which a polynucleotide is transcribed into mRNA and/or the process by which the transcribed mRNA (also referred to as "transcript") is subsequently being translated into peptides, polypeptides, or proteins. The transcripts and the encoded polypeptides are collectedly referred to as gene product. If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in an eukaryotic cell.

"Differentially expressed", as applied to nucleotide sequence or polypeptide sequence in a subject, refers to over-expression or under-expression of that sequence when compared to that detected in a control. Underexpression also encompasses absence of expression of a particular sequence as evidenced by the absence of detectable expression in a test subject when compared to a control.

"Differential expression" refers to alterations in the abundance or the expression pattern of a gene product.

A "primer" is a short polynucleotide, generally with a free 3' -OH group, that binds to a target or "template" potentially present in a sample of interest by hybridizing with the target, and thereafter promoting polymerization of a polynucleotide complementary to the target.

The term "hybridize" as applied to a polynucleotide refers to the ability of the polynucleotide to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues in a hybridization reaction. The hydrogen bonding may occur by Watson-Crick base pairing, Hoogstein binding, or in any other sequence-specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi-stranded complex, a single self-hybridizing strand, or any combination of these. The hybridization reaction may constitute a step in a more extensive process, such as the initiation of a PCR reaction, or the enzymatic cleavage of a polynucleotide by a ribozyme.

Hybridization can be performed under conditions of different "stringency". Relevant conditions include temperature, ionic strength, time of incubation, the presence of additional solutes in the reaction mixture such as formamide, and the washing procedure. Higher stringency conditions are those conditions, such as higher temperature and lower sodium ion concentration, which require higher minimum complementarity between hybridizing elements for a stable hybridization complex to form. In general, a low stringency hybridization reaction is carried out at about 40 ^{0}C in 10 x SSC or a solution of equivalent ionic strength/temperature. A moderate stringency hybridization is typically performed at about 50 ^{0}C in 6 x SSC, and a high stringency hybridization reaction is generally performed at about 60 ^{0}C in 1 x SSC.

When hybridization occurs in an antiparallel configuration between two single-stranded polynucleotides, the reaction is called "annealing" and those polynucleotides are described as "complementary". A double-stranded polynucleotide can be "complementary" or "homologous" to another polynucleotide, if hybridization can occur between one of the strands of the first polynucleotide and the second. "Complementarity" or "homology" (the degree that one polynucleotide is complementary with another) is quantifiable in terms of the proportion of bases in opposing strands that are expected to form hydrogen bonding with each other, according to generally accepted base-pairing rules.

In the context of polynucleotides, a "linear sequence" or a "sequence" is an order of nucleotides in a polynucleotide in a 5' to 3' direction in which residues that neighbor each other in the sequence are contiguous in the primary structure of the

polynucleotide. A "partial sequence" is a linear sequence of part of a polynucleotide which is known to comprise additional residues in one or both directions.

The terms "cytosolic", "nuclear" and "secreted" as applied to cellular proteins specify the extracellular and/or subcellular location in which the cellular protein is mostly localized. Certain proteins are "chaperons", capable of translocating back and forth between the cytosol and the nucleus of a cell.

A "subject" as used herein refers to a biological entity containing expressed genetic materials. The biological entity is preferably can be plant, animal, or microorganisms including bacteria, viruses, fungi, and protozoa. Tissues, cells and their progeny of a biological entity obtained *in vivo* or cultured *in vitro* are also encompassed.

A "control" is an alternative subject or sample used in an experiment for comparison purpose. A control can be "positive" or "negative". For example, where the purpose of the experiment is to detect a differentially expressed transcript or polypeptide in cell or tissue affected by a disease of concern, it is generally preferable to use a positive control (a subject or a sample from a subject, exhibiting such differential expression and syndromes characteristic of that disease), and a negative control (a subject or a sample from a subject lacking the differential expression and clinical syndrome of that disease).

"Heterologous" means derived from a genotypically distinct entity from the rest of the entity to which it is being compared. For example, a promoter removed from its native coding sequence and operatively linked to a coding sequence other than the native sequence is a heterologous promoter.

A "cell line" or "cell culture" denotes bacterial, plant, insect or higher eukaryotic cells grown or maintained in vitro. The descendants of a cell may not be completely identical (either morphologically, genotypically, or phenotypically) to the parent cell.

A "vector" is a nucleic acid molecule, preferably self-replicating, which transfers an inserted nucleic acid molecule into and/or between host cells. The term includes vectors that function primarily for insertion of a DNA or RNA into a cell, replication of vectors that function primarily for the replication of DNA or RNA, and expression vectors that function for transcription and/or translation of the DNA

or RNA. Also included are vectors that provide more than one of the above functions.

An "expression vector" is a polynucleotide which, when introduced into an appropriate host cell, can be transcribed and translated into a polypeptide(s). An "expression system" usually connotes a suitable host cell comprised of an expression vector that can function to yield a desired expression product.

A "replicon" refers to a polynucleotide comprising an origin of replication (generally referred to as an ori sequence) which allows for replication of the polynucleotide in an appropriate host cell. Examples of replicons include episomes (such as plasmids), as well as chromosomes (such as the nuclear or mitochondrial chromosomes).

A "transcription unit" is a DNA segment capable of directing transcription of a gene or fragment thereof. Typically, a transcription unit comprises a promoter operably linked to a gene or a DNA fragment that is to be transcribed, and optionally regulatory sequences located either upstream or downstream of the initiation site or the termination site of the transcribed gene or fragment.

Vectors of the present invention

A central aspect of the present invention is the design of a recombinant vector suited for bi-directional transcription of a transgene to yield both sense and antisense RNA transcripts of the transgene in a eukaryotic cell. The invention vectors are particularly suited for mediating nuclear gene silencing in a variety of biological systems. Distinguished from the previously described DNA vectors, the subject vectors have the following unique characteristics: (a) the vector replicates and directs expression of a transgene in a eukaryotic cell; and (b) the vector comprises a replicon having two overlapping transcription units arranged in an opposing orientation and flanking a transgene of interest, wherein the two overlapping transcription units yield both sense and antisense RNA transcripts from the same transgene in a eukaryotic host cell.

Several factors apply to the design of vectors having the above-mentioned characteristics. First, the vector comprises a replicon having an origin of replication (generally referred to as an ori sequence) which permits replication of the vector in a eukaryotic host cell. A preferred replicon is one comprising viral sequences capable

of directing autonomous replication of the vector in an appropriate host cell. Non-limiting examples of viral replicons include sequences derived from DNA viruses such as *Geminivirus, Caulimoviridae, Badnaviridae; Circoviridae, Circinoviridae, Parvoviridae, Papovaviridae, Polyomaviridae, Adenoviridae, Herpesviridae, Poxviridae, Iridoviridae, Baculoviridae, Hepadnaviridae, Retroviridae, Gyrovirus, Nanovirus*, and African Swine Fever virus, or the like. In addition to the replication origin, a replicon typically carries a transcription unit that directs transcription of a transgene or a fragment thereof to yield a plurality of RNA transcripts.

A second consideration in designing the subject vector is to select two overlapping transcription units. By "overlapping" is meant that the two transcription units directs transcription of both DNA strands of the same transgene to yield a plurality of partially or perfectly double stranded RNA transcripts. The two overlapping transcription units are typically arranged in an opposing orientation so that each unit can drive transcription of one of the complementary strands from the same transgene, and thus facilitate the generation of double stranded RNA transcripts. Elements within a transcription unit include but are not limited to promoter regions, enhancer regions, repressor binding regions, transcription initiation sites, ribosome binding sites, translation initiation sites, protein encoding regions and introns, and termination sites for transcription and translation. Preferred transcription units are arranged in a configuration shown in Figure 2(a)-(d).

As used herein, a "promoter" is a DNA region capable under certain conditions of binding RNA polymerase and initiating transcription of a coding region located downstream (in the 3' direction) from the promoter. It can be constitutive or inducible. In general, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence is a transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes.

The choice of promoters will largely depend on the host cells in which the vector is introduced. Commonly employed plant promoters include but are not limited those from agrobacterium, nopaline synthase gene, octopine synthase gene,

mannopine synthase, rbcS (small subunit of ribulose bis-phosphate carboxylase). In addition, the promoter sequences may be provided by viral material. Any RNA virus subgenomic promoters described in Dawson et al. Advances in Virus Research, **38**:307-342 and WO93/03161 can thus be employed. For animal cells, a variety of robust promoters, both viral and non-viral promoters, are known in the art. Non-limiting representative viral promoters include CMV, the early and late promoters of SV40 virus, promoters of various types of adenoviruses (e.g. adenovirus 2) and adeno-associated viruses. It is also possible, and often desirable, to utilize promoters normally associated with a desired transgene sequence, provided that such control sequences are compatible with the host cell system. See Goeddel et al., Gene Expression Technology Methods in Enzymology Volume 185, Academic Press, San Diego, (1991), Ausubel et al, Protocols in Molecular Biology, Wiley Interscience (1994).

Suitable promoter sequences for other eukaryotic cells such as yeast cells include the promoters for 3-phosphoglycerate kinase, or other glycolytic enzymes, such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. Other promoters, which have the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization.

To optimize the yield of double-stranded RNAs formed from the sense and anti-sense strands transcribed by the overlapping units, it is preferable to use two promoters of comparable strength. The relative strength of the promoters can be determined or ascertained by any convention recombinant techniques and methods exemplified herein. Representative techniques are Northern blot hybridization and DNA array-based technologies. An illustrative promoter pair comprises MSV mp promoter and CaMV 35S RNA promoter.

Where desired, heterologous promoters that are removed from their native coding sequences and operatively linked to a transgene which it is not naturally

found linked, can be used in constructing the invention vectors. As such, any viral promoters described above can be used to drive the transcription of a non-viral transgenes; promoters of one class of genes can be employed to direct transcription of transgenes coding for other related or unrelated classes of proteins. In certain embodiments of the invention, it is preferable to employ inducible promoters to control the transcription of a transgene. A diverse variety of inducible promoters have been described in the art. Promoters of any endogenous genes whose expressions are inducible by internal or external factors can be employed. Factors applicable for transcription induction include but are not limited to hormones, heat shock, oxygen deficiency, light, stress and various chemicals. Commonly employed inducible promoters are β-gal promoter that is activated upon addition of IPTG; hps70 promoter that is inducible by heat shock; and ribulose-1,5-biphosphate carboxylase (RUBISCO) promoter that is regulated by light.

Tissue-specific promoters may also be used. A vast diversity of tissue specific promoters have been described and employed by artisans in the field. Representative plant tissue promoters include that of legumin (or other seed storage protein promoters), patatin and the like. Exemplary promoters operative in selective animal tissue include hepatocyte-specific promoters and cardiac muscle specific promoters. Depending on the intended use of the subject vectors, those skilled in the art will know of other suitable tissue-specific promoters applicable for non-constitutive bi-directional transcription.

In constructing the subject vectors, the termination sequences associated with the transgene are also inserted into the 3' end of the sequence desired to be transcribed to provide polyadenylation of the mRNA and/or transcriptional termination signal. The terminator sequence preferably contains one or more transcriptional termination sequences (such as polyadenylation sequences) and may also be lengthened by the inclusion of additional DNA sequence so as to further disrupt transcriptional read-through. Preferred terminator sequences (or termination sites) of the present invention have a gene that is followed by a transcription termination sequence, either its own termination sequence or a heterologous termination sequence. Examples of such termination sequences, including stop codons coupled to various polyadenylation sequences that are known in the art, widely available, and exemplified below. Where the terminator comprises a gene, it

can be advantageous to use a gene which encodes a detectable or selectable marker; thereby providing a means by which the presence and/or absence of the terminator sequence (and therefore the corresponding inactivation and/or activation of the transcription unit) can be detected and/or selected. Alternatively, a terminator may simply be a second promoter, arranged in inverted orientation to the promoter described above.

The terminators and promoters of the two overlapping transcription units may take a variety of configurations. In one aspect, terminators 1 and 2 of the overlapping transcription units are arranged to immediately flank the transgene as shown in Figure 2(a). In another aspect, the two terminators are placed at the 5' end or the 3' end of their respective promoters as depicted in Figure 2(b). In other aspects, terminator 1 and promoter 1 are flanked by terminator 2 and promoter 2 as shown in Figure 2(c), or vice versa (see Figure 2(d)). Any other variations in configuring the two overlapping transcription units that permit bi-directional transcription are encompassed by the present invention.

The transgene transcribed by an invention vector can be any gene expressed in a eukaryotic cell. The selection of transgene is determined largely by the intended purpose of the vector. Where the vector is used to inhibit expression of an endogenous gene present in a host cell, the transgene selected are substantially homologous to the target endogenous gene. In general, substantially homologous nucleotide sequences are at least about 60% identical with each other, after alignment of the homologous regions. Preferably, the sequences are at least about 75% identical; more preferably, they are at least about 80% identical; more preferably, they are at least about 90% identical; still more preferably, the sequences are 95% identical.

Sequence alignment and homology searches are often determined with the aid of computer methods. A variety of software programs are available in the art. Non-limiting examples of these programs are Blast (http://www.ncbi.nlm.nih.gov/BLAST/), Fasta (Genetics Computing Group package, Madison, Wisconsin), DNA Star, MegAlign, and GeneJocky. Any sequence databases that contains DNA sequences corresponding to a target gene or a segment thereof can be used for sequence analysis. Commonly employed databases include but are not limited to GenBank, EMBL, DDBJ, PDB, SWISS-PROT, EST,

STS, GSS, and HTGS. Sequence similarity can be discerned by aligning the transgene sequence against a target endogenous gene sequence. Common parameters for determining the extent of homology set forth by one or more of the aforementioned alignment programs include p value and percent sequence identity. P value is the probability that the alignment is produced by chance. For a single alignment, the p value can be calculated according to Karlin et al. (1990) *Prco.Natl. Acad. Sci* **87**: 2264. For multiple alignments, the p value can be calculated using a heuristic approach such as the one programmed in Blast. Percent sequence identity is defined by the ratio of the number of nucleotide matches between the query sequence and the known sequence when the two are optimally aligned. A selected transgene and target endogenous sequences are considered to be substantially homologous when the regions of alignment exhibit the aforementioned range of percentage of identity using Fasta or Blast alignment program with the default settings.

Sequence homology can also be determined by functional analyses. A sequence that preserves the functionality of the nucleic acid with which it is being compared is particularly preferred. Functionality may be established by different criteria, such as ability to hybridize with a target polynucleotide, ability to effectively amplify a target sequence to yield a substantially homogenous multiplicity of products, and the ability to extend the 3' end sequence complementary to a target sequence in a nucleotide sequencing reaction.

Where desired, the transgene may comprise heterologous sequences that facilitate detection of the expression and purification of the gene product. Examples of such sequences are known in the art and include those encoding reporter proteins such as β-galactosidase, β -lactamase, chloramphenicol acetyltransferase (CAT), luciferase, green fluorescent protein (GFP) and their derivatives. Other heterologous sequences that facilitate purification may code for epitopes such as Myc, HA (derived from influenza virus hemagglutinin), His-6, FLAG, glutathione S-transferase (GST), maltose-binding protein (MBP), or the Fc portion of immunoglobulin.

The target endogenous genes whose expression is to be inhibited encompass native and heterologous genes present in the host cell. "Native" genes are nucleic acid sequences originated from the host cell. Non-limiting illustrative native genes

include those encode membrane proteins, cytosolic proteins, secreted proteins, nuclear proteins and chaperon proteins. Heterologous genes are sequences acquired exogenously by the host cell. Exogenous sequences can be either integrated into the host cell genome, or maintained as episomal sequences. An exemplary class of heterologous genes includes pathogenic genes derived from viruses, bacteria, fungi, and protozoa.

The endogenous genes suitable for the present invention may also be characterized based on one or more of the following features: ability to induce a phenotypic change in a host cell or organism, species origin, developmental origin, primary structural similarity, involvement in a particular biological process, association with or resistance to a particular disease or disease stage, tissue, sub-tissue or cell-specific expression pattern, and subcellular location of the expressed gene product. In one aspect, the endogenous gene may be any gene expressed in a eukaryote cell, such as a plant cell, animal cell or a yeast cell. In another aspect, the endogenous gene confers a phenotypic characteristic detectable by visual, microscopic, genetic, or chemical means. Within this class of genes, of particular interest are plant genes involved in growth phenotypes, e.g. stunting, hyperbranching, vein banding, ring spot, etching, and those responsible for color characteristics including bleaching and chlorosis. Also, of particular relevance are genes which upon inhibition provide an enhanced resistance to pathogens (e.g. bacteria, fungi, viruses, insects, and protozoa), and resistance to adverse environmental factors (e.g. temperature fluctuation, nutritional deficiency, adverse soil conditions, moisture, dryness, etc.).

In another aspect, the endogenous genes are of a specific developmental origin, such as those expressed in an embryo or an adult organism, during ectoderm, mesoderm, or endoderm formation in a multi-cellular animal, or during development of leaves, tubers, bud of a plant. In yet another aspect, the endogenous genes belong to a family of genes, or a sub-family of genes that share primary structural similarities. Structural similarities can be discerned with the aid of computer software described above. Non-limiting examples of gene families include those encoding proteinase, proteinase inhibitors, cell surface receptors, protein kinases (e.g. tyrosine, serine/threonine or histidine kinases), trimeric G-proteins, cytokines, PH-, SH2-, SH3-, PDZ-domain containing proteins, and any of those gene families

published by the Institute for Genomic Research (TIGR), Incyte Pharmaceuticals, Inc., Human Genome Sciences Inc., Monsanto, and PE-Celera.

In yet another aspect, the endogenous genes are involved in a specific biological process, including but not limited to cell cycle regulation, cell differentiation, chemotaxsis, apoptosis, cell motility and cytoskeletal rearrangement. In still another aspect, the endogenous genes embodied in the invention are associated with a particular disease or with a specific disease stage. Such genes include but are not limited to those associated with autoimmune diseases, obesity, hypertension, diabetes, neuronal and/or muscular degenerative diseases, cardiac diseases, endocrine disorders, any combinations thereof. In yet still another aspect, the endogenous genes encompass those exhibiting restricted expression patterns. Non-limiting exemplary gene transcripts of this class include those that are not ubiquitously expressed, but rather are differentially expressed in one or more of the plant tissues including leaf, seed, tuber, stems, root, and bud; or expressed in animal body tissues including heart, liver, prostate, lung, kidney, bone marrow, blood, skin, bladder, brain, muscles, nerves, and selected tissues that are affected by various types of cancer (malignant or non-metastatic), affected by cystic fibrosis or polycystic kidney disease. Additional examples of non-ubiquitously expressed genes are those whose gene products are localized to certain subcellular locations: extracellular matrix, nucleus, cytoplasm, cytoskeleton, plasma and/or intracellular membranous structures which include but are not limited to coated pits, Golgi apparatus, endoplasmic reticulum, endosome, lysosome, and mitochondria.

In addition to the above-described elements, the vectors may contain a selectable marker (for example, a gene encoding a protein necessary for the survival or growth of a host cell transformed with the vector), although such a marker gene can be carried on another polynucleotide sequence co-introduced into the host cell. Only those host cells into which a selectable gene has been introduced will survive and/or grow under selective conditions. Typical selection genes encode protein(s) that (a) confer resistance to antibiotics or other toxins substances, e.g., ampicillin, neomycyin, methotrexate, etc.; (b) complement auxotrophic deficiencies; or (c) supply critical nutrients not available from complex media. The choice of the proper marker gene will depend on the host cell, and appropriate genes for different hosts are known in the art.

The vectors embodied in this invention can be obtained using recombinant cloning methods and/or by chemical synthesis. A vast number of recombinant cloning techniques such as PCR, restriction endonuclease digestion and ligation are well known in the art, and need not be described in detail herein. One of skill in the art can also use the sequence data provided herein or that in the public or proprietary databases to obtain a desired vector by any synthetic means available in the art.

Host cell and transgenic organisms of the present invention:

The invention provides eukaryotic host cells transformed with the recombinant DNA vectors described above. The recombinant vectors containing the transgene of interest can be introduced into a suitable eukaryotic cell by any of a number of appropriate means, including electroporation, transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; and infection (where the vector is coupled to an infectious agent). The choice of introducing vectors will often depend on features of the host cell.

For most animal cells, any of the above-mentioned methods is suitable for vector delivery. For plant cells, a variety of techniques derived from these general methods is available in the art. The host cells may be in the form of whole plants, isolated cells or protoplasts. Preferably, the cells are "intact" in that the cell comprises an outer layer of cell wall, typically composed of cellulose for protection and maintaining the rigidity of the plant cell. Illustrative procedures for introducing vectors into plant cells include Agrobacterium-mediated plant transformation, protoplast transformation, gene transfer into pollen, injection into reproductive organs and injection into immature embryos. As is evident to one skilled in the art, each of these methods has distinct advantages and disadvantages. Thus, one particular method of introducing genes into a particular plant species may not necessarily be the most effective for another plant species.

Agrobacterium tumefaciens-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, bypassing the need for regeneration of an intact plant from a protoplast. The use of Agrobacterium-mediated expression vectors to introduce

DNA into plant cells is well known in the art. This technique makes use of a common feature of *Agrobacterium* which colonizes plants by transferring a portion of their DNA (the T-DNA) into a host cell, where it becomes integrated into nuclear DNA. The T-DNA is defined by border sequences which are 25 base pairs long, and any DNA between these border sequences is transferred to the plant cells as well. The insertion of a recombinant plant viral nucleic acid between the T-DNA border sequences results in transfer of the recombinant plant viral nucleic acid to the plant cells, where the recombinant plant viral nucleic acid is replicated, and then spreads systemically through the plant. Agro-infection has been accomplished with potato spindle tuber viroid (PSTV); CaV; and Lazarowitz, S., *Nucl. Acids Res.* 16:229 (1988)) digitaria streak virus (Donson *et al., Virology* 162:248 (1988)), wheat dwarf and tomato golden mosaic virus (TGMV). Therefore, agro-infection of a susceptible plant could be accomplished with a virion containing a recombinant plant viral nucleic acid based on the nucleotide sequence of any of the above viruses. Particle bombardment or electrosporation or any other methods known in the art may also be used.

Because not all plants are natural hosts for Agrobacterium, alternative methods such as transformation of protoplasts may be employed to introduce the subject vectors into the host cells. For certain monocots, transformation of the plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments. See, for example, Potrykus et al., *Mol. Gen. Genet.*, **199**:167-177 (1985); Fromm et al., *Nature*, **319**:791 (1986); Callis et al., *Genes and Development*, **1**:1183 (1987). Applicability of these techniques to different plant species may depend upon the feasibility to regenerate that particular plant species from protoplasts.

In addition to protoplast transformation, particle bombardment is an alternative and convenient technique for delivering the invention vectors into a plant host cell. Specifically, the plant cells may be bombarded with microparticles coated with a plurality of the subject vectors. Bombardment with DNA-coated microprojectiles has been successfully used to produce stable transformants in both plants and animals (see, for example, Sanford et al. (1993) *Methods in Enzymology*, **217**:483-509). Microparticles suitable for introducing vectors into a plant cell are

typically made of metal, preferably tungsten or gold. These microparticles are available for example, from BioRad (*e.g.,* Bio-Rad's PDS-1000/He). Those skilled in the art will know that the particle bombardment protocol can be optimized for any plant by varying parameters such as He pressure, quantity of coated particles, distance between the macrocarrier and the stopping screen and flying distance from the stopping screen to the target.

Vectors can also be introduced into plants by direct DNA transfer into pollen as described by Zhou et al., Methods in Enzymology, **101**:433 (1983); Luo et al., *Plant Mol. Biol. Reporter*, **6**:165 (1988). Alternatively, the vectors can be injected into reproductive organs of a plant as described by Pena et al., *Nature*, **325**:274 (1987).

Other techniques for introducing nucleic acids into a plant cell include:

(a) Hand Inoculations. Hand inoculations are performed using a neutral pH, low molarity phosphate buffer, with the addition of celite or carborundum (usually about 1%). One to four drops of the preparation is put onto the upper surface of a leaf and gently rubbed.

(b) Mechanized Inoculations of Plant Beds. Plant bed inoculations are performed by spraying (gas-propelled) the vector solution into a tractor-driven mower while cutting the leaves. Alternatively, the plant bed is mowed and the vector solution sprayed immediately onto the cut leaves.

(c) High Pressure Spray of Single Leaves. Single plant inoculations can also be performed by spraying the leaves with a narrow, directed spray (50 psi, 6-12 inches from the leaf) containing approximately 1% carborundum in the buffered vector solution.

(d) Vacuum Infiltration. Inoculations may be accomplished by subjecting a host organism to a substantially vacuum pressure environment in order to facilitate infection.

Once introduced into a suitable host cell, expression of the transgene can be determined using any assay known in the art. For example, the presence of transcribed sense or anti-sense strands of the transgene can be detected and/or quantified by conventional hybridization assays (e.g. Northern blot analysis), amplification procedures (e.g. RT-PCR), SAGE (U.S. Patent No. 5,695,937), and

array-based technologies (see e.g. U.S. Pat. Nos. 5,405,783, 5,412,087 and 5,445,934). In conducting these analytical procedures, it is preferable to induce transcription of one strand of the transgene at a time. As is apparent to one skilled in the art, the simultaneous transcription of both sense and anti-sense strands facilitates formation of double stranded RNA molecules, which may obscure the accurate determination of the levels of sense and anti-sense RNA transcripts.

Expression of the transgene can also be determined by examining the protein product. A variety of techniques are available in the art for protein analysis. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using *e.g.*, colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunoflourescent assays, and PAGE-SDS.

In general, determining the protein level involves (a) providing a biological sample containing polypeptides; and (b) measuring the amount of any immunospecific binding that occurs between an antibody reactive to the transgene product and a component in the sample, in which the amount of immunospecific binding indicates the level of expressed proteins. Antibodies that specifically recognize and bind to the protein products of the transgene are required for immunoassays. These may be purchased from commercial vendors or generated and screened using methods well known in the art. See Harlow and Lane (1988) *supra.* and Sambrook et al. (1989) *supra.* The sample of test proteins can be prepared by homogenizing the eukaryotic transformants (e.g. plant cells) or their progenies made therefrom, and optionally solubilizing the test protein using detergents, preferably non-reducing detergents such as triton and digitonin. The binding reaction in which the test proteins are allowed to interact with the detecting antibodies may be performed in solution, or on a solid tissue sample, for example, using tissue sections or solid support that has been immobilized with the test proteins. The formation of the complex can be detected by a number of techniques known in the art. For example, the antibodies may be supplied with a label and unreacted antibodies may be removed from the complex; the amount of remaining label thereby indicating the amount of complex formed. Results obtained using any such assay on a sample

from a plant transformant or a progeny thereof is compared with those from a non-transformed source as a control.

The eukaryotic host cells of this invention are grown under favorable conditions to effect transcription of the transgene. Non-limiting examples of eukaryotic hosts are fungus, yeast, plant cells, insect, avian, mammalian or other animal cells. The host cells can be used, *inter alia*, as repositories of the transgene and/or vehicles for production of the transgene-specific double stranded RNAs. The host cells may also be employed to generate transgenic organisms such as transgenic animals and plants comprising the recombinant DNA vectors of the present invention. Preferred host cells are those having the propensity to regenerate into tissue or a whole organisms. Examples of these preferred host cells are oocytes, blastocytes, and certain plant cells exemplified herein.

Accordingly, this invention provides transgenic plants carrying the subject vectors. In a preferred embodiment, the trangenic plant exhibits a reduced expression (when compared to a control plant) of an endogenous gene that is substantially homologous to the transgene carried in the subject vector.

The regeneration of plants from either single plant protoplasts or various explants is well known in the art. See, for example, Methods for Plant Molecular Biology, Mary A. Shuler and Raymond E. Zielinski, Academic Press, Inc., San Diego, Calif. (1988). This regeneration and growth process includes the steps of selection of transformant cells and shoots, rooting the transformant shoots and growth of the plantlets in soil.

The regeneration of plants containing the subject vector introduced by Agrobacterium tumefaciens from leaf explants can be achieved as described by Horsch et al., *Science*, **227**:1229-1231 (1985). In this procedure, transformants are grown in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant species being transformed as described by Fraley et al., *Proc. Natl. Acad. Sci. U.S.A.*, **80**:4803 (1983). This procedure typically produces shoots within two to four weeks and these transformant shoots are then transferred to an appropriate root-inducing medium containing the selective agent and an antibiotic to prevent bacterial growth. Transformant shoots that rooted in the presence of the selective agent to form plantlets are then transplanted to soil to allow

the production of roots. These procedures will vary depending upon the particular plant species employed, as is apparent to one of ordinary skill in the art.

A population of progeny can be produced from the first and second transformants of a plant species by methods well known in the art including cross fertilization and asexual reproduction. Transgenic plants embodied in the present invention are useful for production of desired proteins, and as test systems for analysis of the biological functions of a gene.

Uses of the vectors of the present invention:

The subject vectors provide specific reagents for inhibiting expression of an endogenous gene present in a host cell. The expression inhibition methods may be used in a wide variety of circumstances including suppression of a gene associated with a particular disease or disease stage; delineating the biological functions of a gene by analyzing a phenotypic change in the host cell that correlates with the selective suppression of gene expression; and facilitating drug screening by rendering the host cell more susceptible or resistant to a therapeutic agent of interest.

Accordingly, this invention provides a method of inhibiting expression of an endogenous gene present in a eukaryotic cell. The method comprises the steps of: (a) providing a subject vector containing a transgene that is substantially homologous to an endogenous gene of a eukaryotic cell; (b) introducing the recombinant vector into the eukaryotic cell; (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene, and thereby inhibiting expression of the corresponding endogenous gene in the eukaryotic cell.

In a separate embodiment, the invention provides a method of identifying a biological function(s) of an endogenous gene of interest in a eukaryotic cell by selectively inhibiting the expression of the endogenous gene. The method involves: (a) providing a recombinant vector of the present invention, wherein the transgene contained in the vector is substantially homologous to the endogenous gene; (b) introducing the recombinant vector of (a) into the eukaryotic cell; (c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene contained in the recombinant vector and thereby inhibiting expression of the endogenous gene in the eukaryotic cell; and

(d) determining one or more phenotypic changes in the eukaryotic cell that correlate with the inhibited expression of the endogenous gene, thereby identifying the biological function(s) of the endogenous gene in the eukaryotic cell.

The host cells encompassed by these embodiments are eukaryotic cells susceptible to dsRNA-mediated "genetic interference". dsRNA induced gene silencing has been observed in a variety of multi-cellular organisms including but not limited to worms, fruitflies, protozoa, fungi, mammals, and zebrafish. Thus, cells from any of these exemplary organisms can be employed. Suitable host cells may be derived from primary cultures or subcultures generated by expansion and/or cloning of primary cultures. Any cells capable of growth in culture can be used as host cells. Of particular interest is the type of cell that differentially expresses (over-expresses or under-expresses) a disease-causing gene. As is apparent to one skilled in the art, various cell lines may be obtained from public or private repositories. The largest depository agent is American Type Culture Collection (http://www.atcc.org), which offers a diverse collection of well-characterized cell lines derived from a vast number of organisms and tissue samples.

Upon delivery of the subject vectors, the host cells are cultured under conditions favorable for gene transcription. The parameters governing eukaryotic cell survival are generally applicable for induction of gene transcription. The culture conditions are well established in the art. Physicochemical parameters which may be controlled *in vitro* are, e.g., pH, CO_2, temperature, and osmolarity. The nutritional requirements of cells are usually provided in standard media formulations developed to provide an optimal environment. Nutrients can be divided into several categories: amino acids and their derivatives, carbohydrates, sugars, fatty acids, complex lipids, nucleic acid derivatives and vitamins. Apart from nutrients for maintaining cell metabolism, most cells also require one or more hormones from at least one of the following groups: steroids, prostaglandins, growth factors, pituitary hormones, and peptide hormones to survive or proliferate (Sato, G.H., et al. in "Growth of Cells in Hormonally Defined Media", Cold Spring Harbor Press, N.Y., 1982; Barnes and Sato (1980) *Anal. Biochem.*, **102**:255. Given the vast wealth of information on the nutrient requirements, medium conditions optimized for cell survival, one skilled in the art can readily fashion various culture conditions using

any one of the aforementioned methods and compositions, alone or in any combination.

The inhibition of expression of the endogenous gene sharing substantial sequence homology with the transgene carried in the vectors can be determined by assaying for a difference, between the host cell and the control cell, in the level of mRNA transcripts of the endogenous gene. Alternatively, a suppression in expression is determined by detecting a difference in the level of the polypeptide(s) encoded by the endogenous gene. A preferred method is to detect a phenotypic change resulting from the decrease in expression of the endogenous gene of interest.

In assaying for an alteration in mRNA level, nucleic acid contained in the host cells is first extracted according to standard methods in the art. For instance, mRNA can be isolated using various lytic enzymes or chemical solutions according to the procedures set forth in Sambrook et al. (1989), *supra* or extracted by nucleic-acid-binding resins following the accompanying instructions provided by manufacturers. The mRNA contained in the extracted nucleic acid sample is then detected by hybridization (e.g. Northern blot analysis) and/or amplification procedures according to methods widely known in the art or based on the methods exemplified herein.

Reduction in expression of the endogenous gene can also be determined by examining the protein product of the endogenous gene. A variety of techniques is available in the art for protein analysis. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using *e.g.*, colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunoflourescent assays, and SDS-PAGE. In addition, cell sorting analysis can be employed to detect cell surface antigens. Such analysis involves labeling target cells with antibodies coupled to a detectable agent, and then separating the labeled cells from the unlabeled ones in a cell sorter. A sophisticated cell separation method is fluorescence-activated cell sorting (FACS). Cells traveling in single file in a fine stream are passed through a laser beam, and the fluorescence of each cell bound by the fluorescently labeled antibodies is then measured.

Antibodies that specifically recognize and bind to the protein products of interest are required for conducting the aforementioned protein analyses. These antibodies may be purchased from commercial vendors or generated and screened using methods well known in the art. See Harlow and Lane (1988) *supra.* and Sambrook et al. (1989) *supra.*

Inhibition of gene expression can also result in phenotypic change(s) in a host cell. As used herein, phenotypic change refers to any non-genotypic change that can be detected visually, or analyzed biochemically or genetically. The choice of detection methods will largely depend on the nature of the phenotypic characteristics that are under investigation. For instance, certain phenotypic features of a plant cell can be detected microscopically or macroscopically. These features include improved tolerance to herbicides, improved tolerance to extremes of heat or cold, drought, salinity or osmotic stress; improved resistance to pests (insects, nematodes or arachnids) or diseases (fungal, bacterial or viral), production of enzymes or secondary metabolites; male or female sterility; dwarfness; early maturity; improved yield, vigor, heterosis, nutritional qualities, flavor or processing properties, and the like. Other detectable phenotypic changes are morphological alterations including but not limited to stunting, hyperbranching, vein banding, ring spot, etching, and those responsible for color characteristics including bleaching and chlorosis.

For animal cells, detectable phenotypic changes may encompass alterations in cell cycle regulation, cell differentiation, apoptosis, chemotaxsis, cell motility and cytoskeletal rearrangement. Methods for detecting these phenotypic changes are well-established in the art and hence are not detailed herein.

Other phenotypic changes commonly observed in both plant and animal cells involve differential expression (over-expression or under-expression) of a particular protein due to the selective inhibition of the endogenous gene of interest. Differential gene expression may be analyzed by any chemical means available in the art or those disclosed herein. As is also apparent to artisans, altering expression of one endogenous gene may lead to changes in gene expression profile of a host of genes mapped to the same or related signal transduction pathways. As used herein, "signal transduction" refers to the process by which stimulatory or inhibitory signals are transmitted into and within a cell to elicit an intracellular response. Any

fluctuation in intracellular response of a eukaryotic host cell is also considered as a type of phenotypic change.

Alteration in intracellular response is often determined with the aid of reporter molecules. For example, when examining a signaling cascade involving a fluctuation of intracellular pH condition, pH sensitive molecules such as fluorescent pH dyes can be used as the reporter molecules. In another example where the signaling pathway of a trimeric G_q protein is analyzed, calcium-sensitive fluorescent probes can be employed as reporters. As is apparent to artisans in the field of signal transduction, trimeric G_q protein is involved in a classic signaling pathway, in which activation of G_q stimulates hydrolysis of phosphoinositides by phospholipase C to generate two classes of well-characterized second messengers, namely, diacylglycerol and inositol phosphates. The latter stimulates the mobilization of calcium from intracellular stores, and thus resulting in a transient surge of intracellular calcium concentration, which is a readout measurable with a calcium-sensitive probe.

Another exemplary class of reporter molecules is a reporter gene operably linked to an inducible promoter that can be activated upon the stimulation or inhibition of a signaling pathway. Reporter proteins can also be linked with other proteins whose expression is dependent upon the stimulation or suppression of a given signaling cascade. Commonly employed reporter proteins can be easily detected by a colorimetric or fluorescent assay. Non-limiting examples of such reporter proteins include : β-galactosidase, β -lactamase, chloramphenicol acetyltransferase (CAT), luciferase, green fluorescent protein (GFP) and their derivatives. Those skilled in the art will know of other suitable reporter molecules for assaying changes in a specific signaling transduction readout, or will be able to ascertain such, using routine experimentation.

To discern inhibition of gene expression, one typically conducts a comparative analysis of the subject and appropriate controls. Preferably, a test includes a positive control sample exhibiting a decrease in gene expression and a negative control having an unaltered expression level. The selection of an appropriate control cell or tissue is dependent on the sample cell or tissue initially selected and its phenotype which is under investigation.

In one aspect, the invention methods can be employed to selectively inhibit expression of an endogenous gene that is native to the eukaryotic host cell. Such a gene may encode encodes a protein selected from the group consisting of a membrane protein, a cytosolic protein, a secreted protein, a nuclear protein and a chaperon protein. Of particular interests are endogenous genes that confer phenotypic changes as a result of inhibition of the expression and/or function of the endogenous genes. In another aspect within this embodiment, the endogenous gene is heterologous to the host cell. As used herein, heterologous genes are acquired exogenously by the host cell. Non-limiting examples of heterologous genes are those derived from virus, bacterium, fungus, and protozoa.

In a separate embodiment, the invention methods are used to identify a biological function(s) of an endogenous gene in a eukaryotic cell by examining a phenotypic change associated with the inhibition in its expression and thus loss of biological function. In essence, the subject methods allow the creation of a transient or more long-term gene-specific knock-out system for analyzing the biological function of any endogenous gene of interest.

Kits comprising the vectors of the present invention

The present invention also encompasses kits containing the vectors of this invention in suitable packaging. Kits embodied by this invention include those that allow generation of a double-stranded RNA transcript in a eukaryotic cell.

Each kit necessarily comprises the reagents which render the delivery of vectors into a eukaryotic host cell possible. The selection of reagents that facilitate delivery of the vectors may vary depending on the particular transfection or infection method used. The kits may also contain reagents useful for generating labeled polynucleotide probes or proteinaceous probes for detection of gene silencing. Each reagent can be supplied in a solid form or dissolved/suspended in a liquid buffer suitable for inventory storage, and later for exchange or addition into the reaction medium when the experiment is performed. Suitable packaging is provided. The kit can optionally provide additional components that are useful in the procedure. These optional components include, but are not limited to, buffers, capture reagents, developing reagents, labels, reacting surfaces, means for detection, control samples, instructions, and interpretive information. The kits can be

employed to generate eukaryotic cells whose endogenous genes are selectively inhibited, and transgenic organisms comprising these eukaryotic cells.

Further illustration of the development and use of vectors and assays according to this invention are provided in the Example section below. The examples are provided as a guide to a practitioner of ordinary skill in the art, and are not meant to be limiting in any way.

EXAMPLES

Example 1: Construction of recombinant vectors comprising two opposing transcription units

We have designed a recombinant vector construct useful for silencing nuclear genes in many of the agriculturally-important cereal crops. The vector comprises sequences derived from maize streak geminivirus, isolated MSV-Kom (genbank accession number AF003952, classification: Family *Geminiviridae*, genus *Mastrevirus*, species maize streak virus, designated MSV-Komatipoort. Maize streak virus has a broad host range that encompasses all agriculturally important cereal crops, including but not limited to corn, wheat, rice, barley, rye, sorghum and millet. The methods for construction of infectious geminiviruses are well known to those skilled in the art, and are described in European patent application 8687015.5 as well as in US Patent No. 5,569,597.

We have synthesized a 1618 base pair synthetic DNA that contains the MSV-Kom *repA* and *repB*, long intergenic region (LIR) and short intergenic region (SIR) and thus all sequences that are required for viral replication. Palmer et al.(1999) *Archives of Virology* **144**:1345-1360. This fragment was cloned into the pZeRO-2 vector (Invitrogen) as an *Eco*RI-*Xba*I fragment, to create the plasmid pMSVLSB-1, the sequence of which is shown in Figure 4. A 171 base pair fragment containing the movement protein (mp) promoter of MSV-Kom is synthesised and cloned into the pZeRO-2 vector as an *Hin*dIII-*Eco*RI fragment to create pMSVLSB-2 (sequence shown in Figure 5). The *Apa*I fragment containing the mp promoter is inserted between the two *Apa*I sites in pMSVLSB-1, to create pMSVLSB-3 (sequence shown in Figure 6).

The cauliflower mosaic virus 35S RNA promoter (CaMV 35S promoter) sequence is amplified with a vector containing this sequence (pBI121, from Clontech) as template DNA, using the following PCR primers containing the following restriction sites (shown in italicized): *Eco*RI in CaMV35SF and *Sal*I in CaMV35SR.

CaMV35SF:

TTT*GAATTC*GTCAACATGGTGGAGCAC (SEQ ID NO:1)

CaMV35SR:
TTT*GTCGAC*GTCCTCTCCAAATGAAATGAAC (SEQ ID NO:2)

The CaMV 35S promoter PCR product yielded is digested with *Eco*RI and *Sal*I and the restricted fragments are purified.

The zeocin resistance gene is amplified by PCR with the vector pZeRO-1 (Invitrogen) as template, using the following primers containing the following restriction sites shown in italicized: *Sal*I, *Pac*I and *Not*I in ZeoF and *Xho*I, *Pac*I and *Not*I in ZeoR:

ZeoF:
CCC*GTCGACTTAATTAAGCGGCCGCG*TTTACAATTTCGCCTGATGC (SEQ ID NO:3)

ZeoR:
CCC*CTCGAGTTAATTAAGCGGCCGC*CTCAAAAAGGATCTTCACCTA G (SEQ ID NO:4)

The zeocin resistance gene product yielded is digested with *Xho*I and *Sal*I and purified.

The nopaline synthase (nos) terminator sequence is amplified by PCR with the vector pBI121 (Clontech) as template, using the following primers, with restriction sites *Xho*I in nosF and *Spe*I in nosR italicized:

NosF:
TTT*CTCGAG*CGAATTTCCCCGATCGTTCAAAC (SEQ ID NO:5)

NosR:
TTT*ACTAGT*CCCGATCTAGTAACATAGATGAC (SEQ ID NO:6)

The nos terminator product yielded is digested with *Xho*I and *Spe*I and purified.

The digested CaMV35S promoter, zeocin resistance gene and nos terminator sequences are ligated together with T4 DNA ligase. The ligated product is diluted 1:100 in sterile water and the whole ligation product is re-amplified with the CaMV35SF and nosR primers. The resulting PCR product is digested with *Eco*RI and *Spe*I, purified and ligated with pMSVLSB-3 that is pre-digested with *EcoRI* and *Spe*I. The ligation reaction is used to transform *E. coli* competent cells. Transformants are selected on Luria Agar plates containing both kanamycin (100 μg/ml) and zeocin (50 μg/ml) to select for colonies containing the CaMV35S promoter-zeocin resistance gene-nos terminator cassette inserted into pMSVLSB-3 (Figure 6 and SEQ ID NO:11). Colonies putatively containing the correct plasmid are chosen, plasmid DNA isolated and screened by digestion with *Eco*RI and *Spe*I. One plasmid designated pMSVLSB-4 (Figure 7 and SEQ ID NO:12) is selected.

One of the methods in the art of construction of infectious clones of geminivirus genomes is to clone tandemly duplicated sequences of the geminivirus genome, with at least the LIR duplicated. This allows the virus sequence to escape from the cloning vector *in planta* by a replicative release mechanism. The virus Rep protein is transiently expressed in transfected cells, and induces a nick at each of the stem loop sequences contained within the origin of replication in the LIR. Rolling circle replication is initiated at each nick point, and this results in release of a ssDNA copy of the virus replicon, which is circularized by the Rep protein, and which then replicates autonomously in the plant cell nucleus. The *Xba*I-*Spe*I fragment from pMSVLSB-3, containing the viral LIR and Rep genes is inserted into the unique *Spe*I site in pMSVLSB-4 to create pMSVLSB-5 (Figure 8 and SEQ ID NO:13). The zeocin resistance gene is deleted by digestion with *Not*I; the DNA is recircularized and used to transform *E.coli* to kanamycin resistance with a new vector, pMSVLSB-6 (Figure 9 and SEQ ID NO:14). When the vector is introduced into plant cells, a monomeric copy of the insert is released by replicative release (described above) and replicates autonomously as construct MSVLSB-6 in the nuclei of infected cells.

The restriction map of construct MSVLSB-6 is shown in Figure 3; this genetic construct possesses the following features: (a) the *rep* genes and origins of replication from maize streak geminivirus that are necessary and sufficient for the autonomous replication of the viral construct and its associated foreign DNA in the host plant cell; (b) two overlapping transcription units present in the DNA replicon. The two overlapping transcription units are arranged according to the configuration shown in Figure 2. With reference to Figure 2, "promoter 1" and "terminator 1" in MSVLSB-6 are the MSV mp promoter and transcription termination signals present in the SIR, respectively, and "promoter 2" and "terminator 2" are the CaMV 35S RNA promoter and nos terminator sequences, respectively. The two overlapping transcription units share three unique restriction sites (*Sal*I, *Pac*I and *Not*I) and one non-unique restriction site (*Xho*I) where foreign DNA may be inserted so that it may be transcribed by both promoters to yield at least a partially double stranded RNA duplex of the foreign DNA sequence.

Example 2: Use of recombinant vectors to inhibit or silence gene expression in cereal crops:

Application of pMSVLSB-6 in inhibition of Dwarf1 gene expression in rice

The vector pMSVLSB-6 exemplified above can be employed to inhibit expression of any endogenous gene in a variety of plant host cells. By way of illustration, the rice gene *Dwarf1* is inhibited to duplicate known mutant phenotype using a pMSVLSB-6 containing a fragment of the coding sequence of *Dwarf1* (Genbank accession number AB028602). The gene is amplified from cDNA isolated from rice seedlings. Primer sequences are designed to have homology with the published sequence of Dwarf1. Ashikari *et al.* (1999) *PNAS U.S.A.* **96**:10284-10289. The primer sequences contain *Not*I restriction sites at their 5' ends. The PCR product is digested with *Not*I and cloned into the *Not*I site of pMSVLSB-6 to generate pMSVLSB-6::dwarf1s and pMSVLSB-6::dwarf1a, with the insert cloned in the sense and antisense orientation with respect to the MSV mp promoter, respectively. The *Xba*I-*Spe*I fragment from each of these plasmids is transferred into an *Agrobacterium* binary vector that is commonly used for rice transformation. This vector is used to transform electrocompetent *Agrobacterium* strain LBA4404

(Life Technologies). *Agrobacterium* cultures containing the appropriate plasmids are used in transformation of rice. Transgenic rice is generated by standard protocols (see, e.g. US Patent 5,591,616). The transgenic rice plants display similar phenotypes to the *dwarf*1 mutant described by Ashikari *et al.* (1999) *supra*: they are giberellin-insensitve, dwarfed in comparison with un-silenced transgenic controls, and having broad, dark green leaves, compact pannicles and short, round grains.

Application of pMSVLSB-6 in inhibition of phytoene desaturase expression in maize seedlings

The coding sequence for the maize phytoene desaturase gene (*pds*), having the Genbank accession number U37285, is amplified from cDNA made from RNA isolated from four-day-old maize seedlings, of the cultivar "Golden Cross Bantam". The primers used for amplification of this cDNA have the following sequences containing the *Pac*I sites (italicized) at the 5' ends:

zeapds1330:
TT*TTTAATTA*AGGTCCGCCTGAATTCTCG (SEQ ID NO:7)

zeapds1873
TTT*TTAATTAA*CGGCAAGGCTCACAGTTTG (SEQ ID NO:8)

PCR amplification with these primers and cDNA made from RNA isolated from maize seedlings yields a product of 565 base pairs, which is then digested with *Pac*I. The progenitor plasmid to pMSVLSB-6, pMSVLSB-5 is digested with *Xba*I and *Spe*I to release the MSV and associated overlapping transcription unit sequences from the pZeRO-2 cloning vector as a single 4816 base pair fragment. This fragment is cloned into the *Agrobacterium* binary vector pBin19 (Genbank: U09365) digested with *Xba*I to yield pMSVLSB-7. The plasmid pMSVLSB-7 is digested with *Pac*I and the *pds* PCR fragment is inserted into this position, generating plasmid pMSVLSB-7::*pds*1 (cloned in the sense orientation with respect to the MSV mp promoter) and pMSVLSB-7::*pds*2 (cloned in the antisense orientation with respect to the MSV mp promoter. These two plasmids are each

introduced into *Agrobacterium* strain C58C1(pMP90) (Koncz and Schell, 1985) by electroporation. The *Agrobacterium* containing the binary vector plasmids is grown overnight in Luria Bertani medium containing appropriate selective antibiotics. The bacterial suspension is loaded into a 100 µl Hamilton syringe and injected into three day old maize seedlings (cultivar Golden Cross Bantam) according to methods described by Escudero et al. (1994) in the chapter "Agroinfection" of The Maize Handbook, Freelings M, Walbot V (eds). Plants that are successfully agroinfected display a photobleaching phenotype on the first three leaves, similar to that induced by spraying the plants with the phytoene desaturase-inhibitor norfluorazon.

CLAIMS

What is claimed is:

1. A eukaryotic recombinant vector comprising a viral replicon having two overlapping transcription units arranged in an opposing orientation and flanking a transgene of interest, wherein the two overlapping transcription units yield both sense and antisense RNA transcripts from the same transgene in a eukaryotic host cell.

2. The eukaryotic recombinant vector of claim 1, wherein each of the overlapping transcription units comprises a promoter and a terminator.

3. The eukaryotic recombinant vector of claim 2, wherein the promoter is a constitutive promoter.

4. The eukaryotic recombinant vector of claim 2, wherein the promoter is an inducible promoter.

5. The eukaryotic recombinant vector of claim 2, wherein the promoter is a tissue-specific promoter.

6. The eukaryotic recombinant vector of claim 1, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(a).

7. The eukaryotic recombinant vector of claim 1, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(b).

8. The eukaryotic recombinant vector of claim 1, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(c).

9. The eukaryotic recombinant vector of claim 1, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(d).

10. The eukaryotic recombinant vector of claim 1 that inhibits gene expression of the eukaryotic host cell.

11. The eukaryotic recombinant vector of claim 1, wherein the eukaryotic host cell is selected from the group consisting of fungus, yeast cell, plant cell and animal cell.

12. The eukaryotic recombinant vector of claim 1 that inhibits expression of an endogenous gene present in the host cell, wherein the endogenous gene is substantially homologous to the transgene contained in the overlapping transcription units.

13. The eukaryotic recombinant vector of claim 12, wherein the endogenous gene is native to the host cell.

14. The eukaryotic recombinant vector of claim 12, wherein the endogenous gene is heterologous to the host cell.

15. The eukaryotic recombinant vector of claim 12, wherein the endogenous gene is a pathogenic gene derived from one or more members of the group consisting of virus, bacterium, fungus, and protozoa.

16. The eukaryotic recombinant vector of claim 1, wherein expression of the transgene to yield double-stranded RNA transcripts confers a phenotypic change in the eukaryotic host cell.

17. The eukaryotic recombinant vector of claim 1, wherein the transgene encodes a protein selected from the group consisting of a membrane protein, a cytosolic protein, a secreted protein, a nuclear protein, and a chaperon protein.

18. The eukaryotic recombinant vector of claim 1 that is an autonomously replicating vector.

19. The eukaryotic recombinant vector of claim 1, wherein the viral replicon is derived from a DNA virus.

20. The eukaryotic recombinant vector of claim 19, wherein the DNA virus is selected from the group consisting of *Geminivirus, Caulimoviridae, Badnaviridae; Circoviridae, Circinoviridae, Parvoviridae, Papovaviridae, Polyomaviridae, Adenoviridae, Herpesviridae, Poxviridae, Iridoviridae, Baculoviridae, Hepadnaviridae, Retroviridae, Gyrovirus*, *Nanovirus*, and African Swine Fever virus.

21. A host cell transformed with a vector of claim 1 or 10.

22. The host cell of claim 21 that is a eukaryotic cell selected from the group consisting of fungus, yeast cell, plant cell and animal cell.

23. A transgenic plant comprising a eukaryotic recombinant vector of claim 1 or 10.

24. The transgenic plant of claim 23 exhibiting reduced expression of an endogenous gene that is substantially homologous to the transgene contained in the eukaryotic recombinant vector.

25. A kit for generating a double-stranded RNA transcript in a eukaryotic cell comprising a eukaryotic recombinant vector of claim 1 in suitable packaging.

26. A method of inhibiting expression of an endogenous gene present in a eukaryotic cell, comprising:

(a) providing a eukaryotic recombinant vector of claim 12;

(b) introducing the eukaryotic recombinant vector into the eukaryotic cell;

(c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene that is contained in the transcription units of the vector, and thereby inhibiting expression of the corresponding endogenous gene in the eukaryotic cell.

27. The method of claim 26, wherein the endogenous gene is native to the host cell.

28. The method of claim 26, wherein the endogenous gene is heterologous to the host cell.

29. The method of claim 26, wherein the endogenous gene is a pathogenic gene derived from one or more members of the group consisting of virus, bacterium, fungus, and protozoa.

30. The method of claim 26, wherein inhibition of the endogenous gene confers a phenotypic change in the host cell.

31. The method of claim 26, wherein the host eukaryotic cell is selected from the group consisting of fungus, yeast cell, plant cell, and animal cell.

32. The method of claim 26, wherein the eukaryotic recombinant vector is an autonomously replicating vector.

33. The method of claim 26, wherein the eukaryotic recombinant vector comprises a viral replicon derived from a DNA virus.

34. The method of claim 26, wherein the DNA virus is selected from the group consisting of *Geminivirus, Caulimoviridae, Badnaviridae; Circoviridae, Circinoviridae, Parvoviridae, Papovaviridae, Polyomaviridae, Adenoviridae,*

Herpesviridae, Poxviridae, Iridoviridae, Baculoviridae, Hepadnaviridae, Retrovirida, Gyrovirus, *Nanovirus*, and African Swine Fever virus.

35. The method of claim 26, wherein the eukaryotic recombinant vector comprises two overlapping transcription units, wherein each transcription unit comprises a promoter and a terminator.

36. The method of claim 26, wherein the promoter is a constitutive promoter.

37. The method of claim 26, wherein the promoter is an inducible promoter.

38. The method of claim 26, wherein the promoter is a tissue-specific promoter.

39. The method of claim 35, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(a).

40. The method of claim 35, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(b).

41. The method of claim 35, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(c).

42. The method of claim 35, wherein the promoter and the terminator of the overlapping transcription units are arranged in a configuration shown in Figure 2(d).

43. A method of identifying a biological function(s) of an endogenous gene of interest in a eukaryotic cell by selectively inhibiting the expression of the endogenous gene, the method comprising:

(a) providing a eukaryotic recombinant vector of claim 12;

(b) introducing the eukaryotic recombinant vector of (a) in to the eukaryotic cell;

(c) culturing the eukaryotic cell of (b) under conditions favorable for expression of both sense and antisense RNA transcripts from the transgene contained in the eukaryotic recombinant vector and thereby inhibiting expression of the endogenous gene in the eukaryotic cell; and

(d) determining one or more phenotypic changes in the eukaryotic cell that correlate with the inhibited expression of the endogenous gene, thereby identifying the biological function(s) of the endogenous gene in the eukaryotic cell.

44. The method of claim 43, wherein the eukaryotic cell is selected from the group consisting of fungus, yeast cell, plant cell, and animal cell.

45. The method of claim 43, wherein the eukaryotic cell is a plant cell.

46. The method of claim 43, wherein the eukaryotic cell is an animal cell.

Figure 1

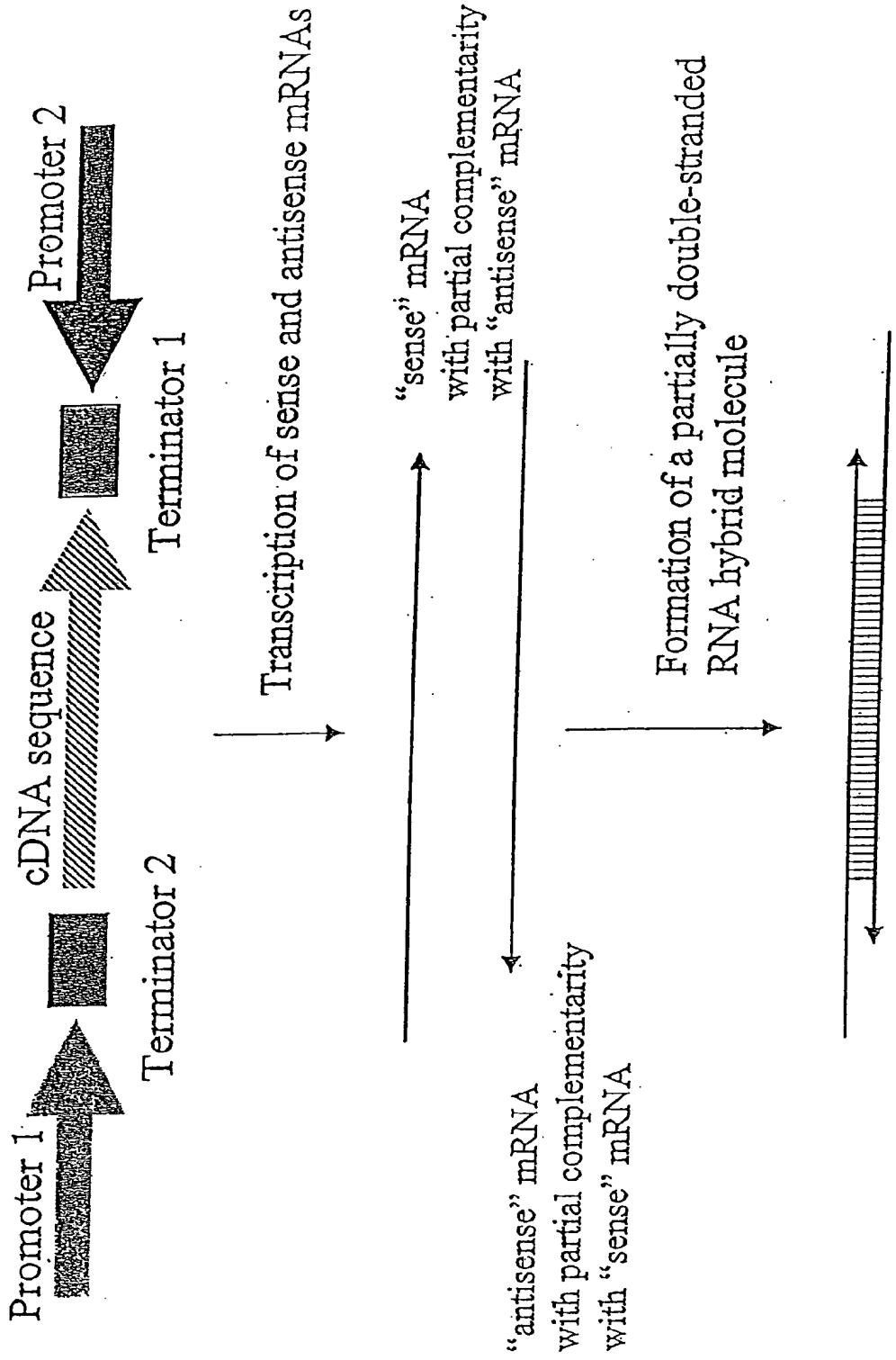
Promoter 1
Terminator 2
cDNA sequence
Terminator 1
Promoter 2
Transcription of sense and antisense mRNAs
"sense" mRNA
with partial complementarity
with "antisense" mRNA
"antisense" mRNA
with partial complementarity
with "sense" mRNA
Formation of a partially double-stranded
RNA hybrid molecule

Figure 2

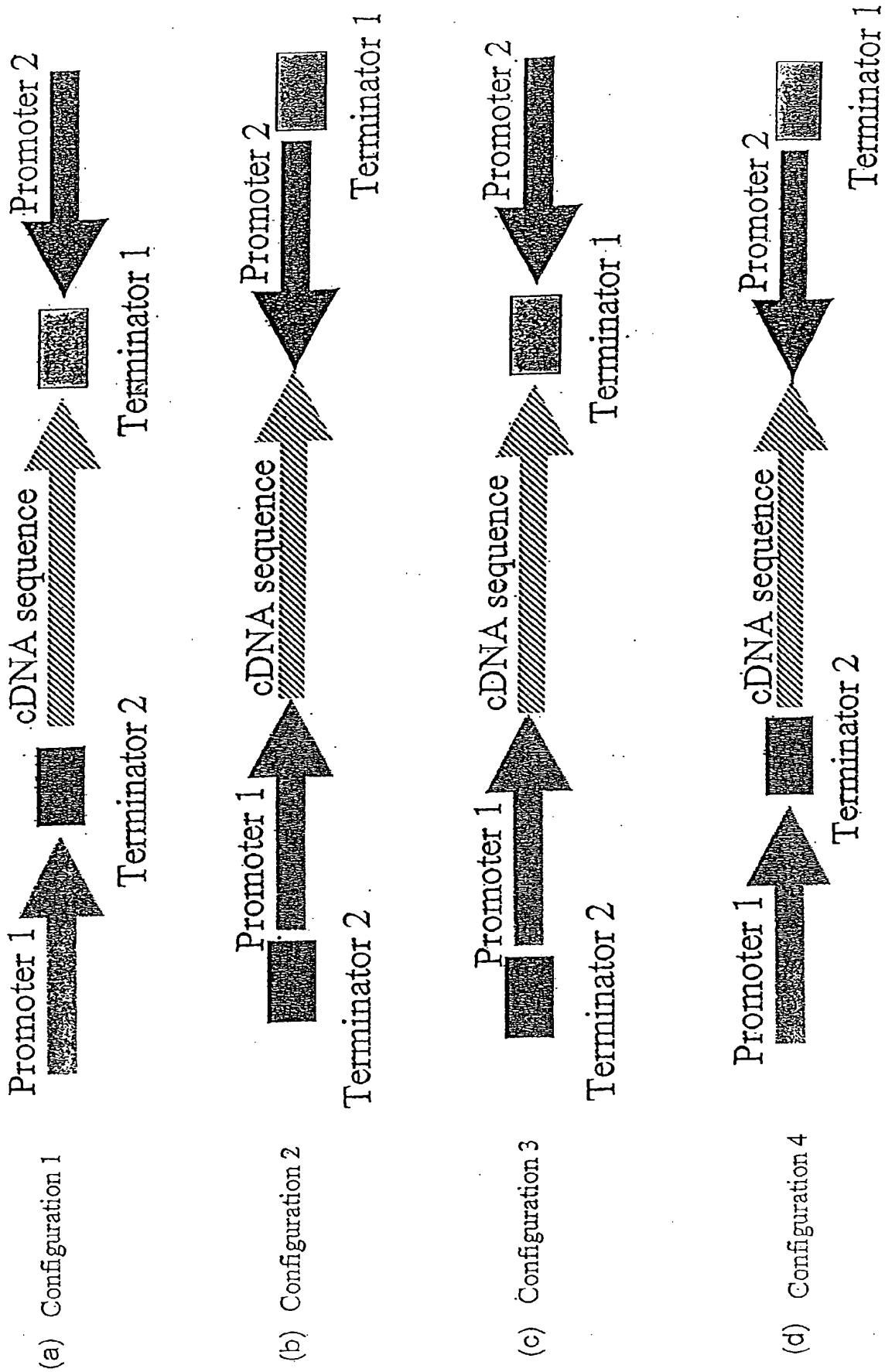
Promoter 1
Terminator 2
cDNA sequence
Promoter 2
Terminator 1
(a) Configuration 1
(b) Configuration 2
(c) Configuration 3
(d) Configuration 4

Figure 3

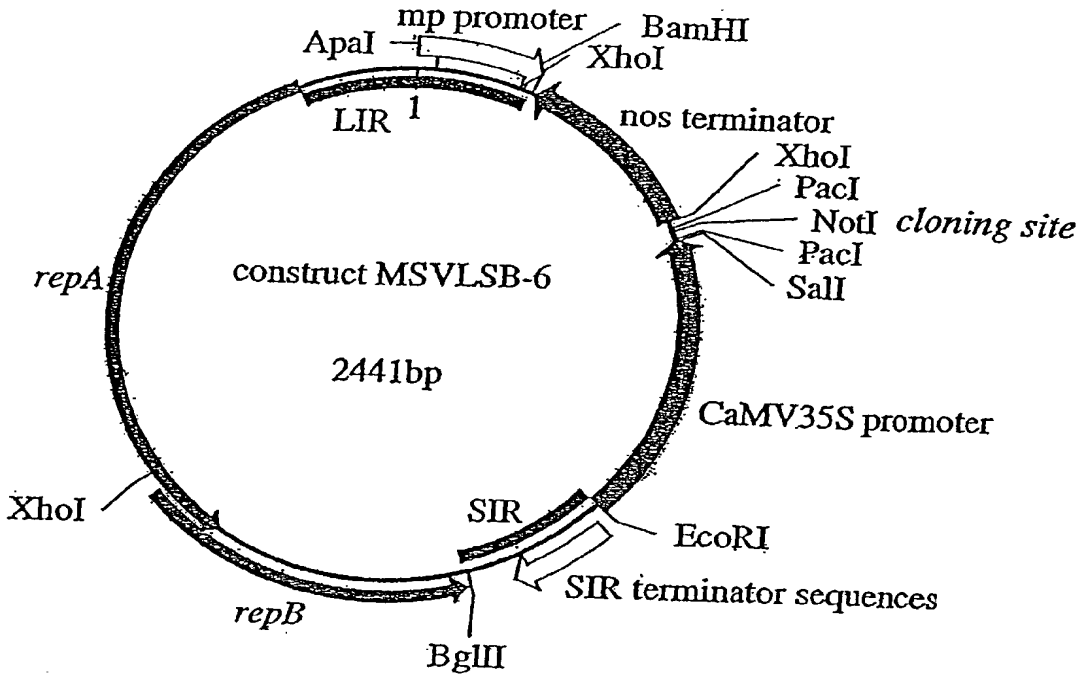
mp promoter
BamHI
ApaI
XhoI
LIR 1
nos terminator
XhoI
PacI
NotI
cloning site
PacI
SalI
repA
construct MSVLSB-6
2441bp
CaMV35S promoter
XhoI
SIR
EcoRI
SIR terminator sequences
repB
BglII

Figure 4

pMSVLSB-1: 4881 bp;
Composition 1161 A; 1260 C; 1251 G; 1209 T; 0 OTHER
Percentage: 24% A; 26% C; 26% G; 25% T; 0%OTHER

Molecular Weight (kDa) ssDNA: 1506.65 dsDNA: 3009.2
ORIGIN

```
1     AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC
61    ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC
121   TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
181   TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT
241   TTAGGTGACA CTATAGAATA CTCAAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA
301   CTAGTAACGG CCGCCAGTGT GCTGGAATTC ATGGGCAGAC CCGTCTGTAC TTTAAGAGTG
361   TTGGCAACCA GTAATGAATA AAAACTCCCG TTTTATTATA TTTGATGAAT GCTGAAAGCT
421   TACATTAATA TGTCGTGCGA TGGCACGAAA AAACACACGC AAACAATACA GGGGGGTAGT
481   CGGCGGGCGG CTAAGGGTGG TGCTCGGCGG GCAGAACATC GAAAAATCAA GATCTATATG
541   AATTACACTT CCTCCGTAGG AGGAAGCACA GGGGGAGAAT ACCACTTCTC CCCCGGCGAC
601   ATAATGTAAA TGACGCAGTT TGCCTCGAAA TACTCCAGCT GCCCTGGAGT CATTTCCTTC
661   ATCCAATCTT CATCCGAGTT GGCGAGGATT ATTGTAGGCT TAGACTTCTT CTGCACCTTT
721   TTCTTCTTAC CATACTTGGG GTTTACAATG AAATCCCTCT GACAGCCAAC TAACTGTTTC
781   CAACAAGGAC AGAATTTAAA CGGAATATCA TCTACGATGT TGTAGATTGC GTCTTCGTTG
841   TATGAAGACC AATCAACATT ATTTTGCCAG TAATTATGAA CCCCTAGGCT TCTGGCCCAA
901   GTAGATTTTC CGGTTCTTGT TGGGCCGACG ATGTAGAGGC TCTGCTTTCT TGATCTTTCA
961   TCTGATGACT GGATACAGAA TCCATCCATT GGAGGTCAGA AATTGCATCC TCGAGGGTAT
1021  AACAGGTAGG TTGAAGGAGC ATGTAAGCTT CGGGACTAAC CTGGAAGATG TTAGGCTGGA
1081  GCCAATCGTT GATTGACTCA TTACAAAGTA AATCAGGTGA GGAGGGTGGA TGAGGATTGG
1141  TGAACTCTTC CTGAATCTCA GGAAAAAGCT TATTTGCAGA GTATTCAAAA TACTGCAATT
1201  TTGTGGACCA ATCAAAGGGG AGCTCTTTCT GGATCATGGA GAGGTACTCT TCTTTGGAGG
1261  TAGCGTGTGA AATAATGTCT CGCATTATTT CATCTTTAGA AGGCTTTTTT TCCTTTACCT
1321  CTGAATCAGA TTTTCCTAGG AAGGGGGACT TCCTAGGAAT GAAAGTACCT CTCTCAAACA
1381  CAGCCAGAGG TTCCTTGAGA ATGTAATCCC TCACTCTGTT AACTGACTTG GCACTCTGAA
1441  TATTTGGGTG AAACCCATTT ATATCAAAGA ACCTTGAGTC AGATATCCTT ATCGGCTTCT
1501  CTGGCTGAAG CAATGCATGT AAATGCAAAC TTCCATCTTT ATGTGCCTCT CGGGCACATA
1561  GAATATATTT GGGAATCCAA CGAACGACGA GCTCCCAGAT CATCTGACAG GCGATTTCAG
1621  GATTTTCTGG ACACTTTGGA TAGGTTAGGA ACGTGTTAGC GTTCCTGTGT GAGAACTGAC
1681  GGTTGGATGA GGAGGAGGCC ATAGCCGACG ACGGAGGTTG AGGCTGAGGG ATGGCAGACT
1741  GGGAGCTCCA AACTCTATAG TATACCCGTG CGCCTTCGAA ATCCGCCGCT CCATTGTCTT
1801  ATAGTGGTTG TAAATGGGCC GGACCGGGCC GGCCCAGCAG GAAAAGAAGG CGCGCACTAA
1861  TATTACCGCG CCTTCTTTTC CTGCGAGGGC CCGGTAGGGA CCGAGCGCTT TGATTTAAAG
1921  CCTGGTTCTG CTTTGCGGCC GCTCGAGCAT GCATCTAGAG GGCCCAATTC GCCCTATAGT
1981  GAGTCGTATT ACAATTCACT GGCCGTCGTT TTACAACGTC GTGACTGGGA AAACCCTGGC
2041  GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGCG TAATAGCGAA
2101  GAGGCCCGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC TATACGTACG GCAGTTTAAG
2161  GTTTACACCT ATAAAAGAGA GAGCCGTTAT CGTCTGTTTG TGGATGTACA GAGTGATATT
2221  ATTGACACGC CGGGGCGACG GATGGTGATC CCCCTGGCCA GTGCACGTCT GCTGTCAGAT
2281  AAAGTCTCCC GTGAACTTTA CCCGGTGGTG CATATCGGGG ATGAAAGCTG GCGCATGATG
2341  ACCACCGATA TGGCCAGTGT GCCGGTCTCC GTTATCGGGG AAGAAGTGGC TGATCTCAGC
2401  CACCGCGAAA ATGACATCAA AAACGCCATT AACCTGATGT TCTGGGGAAT ATAAATGTCA
2461  GGCCTGAATG GCGAATGGAC GCGCCCTGTA GCGGCGCATT AAGCGCGCGG GTGTGGTGGT
2521  TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT TCGCTTTCTT
2581  CCCTTCCTTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC GGGGGCTCCC
2641  TTTAGGGTTC CGATTTAGAG CTTTACGGCA CCTCGACCGC AAAAAACTTG ATTTGGGTGA
2701  TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA CGTTGGAGTC
2761  CACGTTCTTT AATAGTGGAC TCTTGTTCCA AACTGGAACA ACACTCAACC CTATCGCGGT
2821  CTATTCTTTT GATTTATAAG GGATGTTGCC GATTTCGGCC TATTGGTTAA AAAATGAGCT
2881  GATTTAACAA AAATTTTAAC AAAATTCAGA AGAACTCGTC AAGAAGGCGA TAGAAGGCGA
```

Figure 4 (cont'd)

```
2941  TGCGCTGCGA ATCGGGAGCG GCGATACCGT AAAGCACGAG GAAGCGGTCA GCCCATTCGC
3001  CGCCAAGCTC TTCAGCAATA TCACGGGTAG CCAACGCTAT GTCCTGATAG CGGTCCGCCA
3061  CACCCAGCCG GCCACAGTCG ATGAATCCAG AAAAGCGGCC ATTTTCCACC ATGATATTCG
3121  GCAAGCAGGC ATCGCCATGG GTCACGACGA GATCCTCGCC GTCGGGCATG CTCGCCTTGA
3181  GCCTGGCGAA CAGTTCGGCT GGCGCGAGCC CCTGATGCTC TTCGTCCAGA TCATCCTGAT
3241  CGACAAGACC GGCTTCCATC CGAGTACGTG CTCGCTCGAT GCGATGTTTC GCTTGGTGGT
3301  CGAATGGGCA GGTAGCCGGA TCAAGCGTAT GCAGCCGCCG CATTGCATCA GCCATGATGG
3361  ATACTTTCTC GGCAGGAGCA AGGTGAGATG ACAGGAGATC CTGCCCCGGC ACTTCGCCCA
3421  ATAGCAGCCA GTCCCTTCCC GCTTCAGTGA CAACGTCGAG CACAGCTGCG CAAGGAACGC
3481  CCGTCGTGGC CAGCCACGAT AGCCGCGCTG CCTCGTCTTG CAGTTCATTC AGGGCACCGG
3541  ACAGGTCGGT CTTGACAAAA AGAACCGGGC GCCCCTGCGC TGACAGCCGG AACACGGCGG
3601  CATCAGAGCA GCCGATTGTC TGTTGTGCCC AGTCATAGCC GAATAGCCTC TCCACCCAAG
3661  CGGCCGGAGA ACCTGCGTGC AATCCATCTT GTTCAATCAT GCGAAACGAT CCTCATCCTG
3721  TCTCTTGATC AGATCTTGAT CCCCTGCGCC ATCAGATCCT TGGCGGCGAG AAAGCCATCC
3781  AGTTTACTTT GCAGGGCTTC CCAACCTTAC CAGAGGGCGC CCCAGCTGGC AATTCCGGTT
3841  CGCTTGCTGT CCATAAAACC GCCCAGTCTA GCTATCGCCA TGTAAGCCCA CTGCAAGCTA
3901  CCTGCTTTCT CTTTGCGCTT GCGTTTTCCC TTGTCCAGAT AGCCCAGTAG CTGACATTCA
3961  TCCGGGGTCA GCACCGTTTC TGCGGACTGG CTTTCTACGT GAAAAGGATC TAGGTGAAGA
4021  TCCTTTTTGA TAATCTCATG ACCAAAATCC CTTAACGTGA GTTTTCGTTC CACTGAGCGT
4081  CAGACCCCGT AGAAAAGATC AAAGGATCTT CTTGAGATCC TTTTTTTCTG CGCGTAATCT
4141  GCTGCTTGCA AACAAAAAAA CCACCGCTAC CAGCGGTGGT TTGTTTGCCG GATCAAGAGC
4201  TACCAACTCT TTTTCCGAAG GTAACTGGCT TCAGCAGAGC GCAGATACCA AATACTGTCC
4261  TTCTAGTGTA GCCGTAGTTA GGCCACCACT TCAAGAACTC TGTAGCACCG CCTACATACC
4321  TCGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCG TGTCTTACCG
4381  GGTTGGACTC AAGACGATAG TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGGTT
4441  CGTGCACACA GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG
4501  AGCTATGAGA AAGCGCCACG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG
4561  GCAGGGTCGG AACAGGAGAG CGCACGAGGG AGCTTCCAGG GGGAAACGCC TGGTATCTTT
4621  ATAGTCCTGT CGGGTTTCGC CACCTCTGAC TTGAGCGTCG ATTTTTGTGA TGCTCGTCAG
4681  GGGGGCGGAG CCTATGGAAA AACGCCAGCA ACGCGGCCTT TTTACGGTTC CTGGGCTTTT
4741  GCTGGCCTTT TGCTCACATG TTCTTTCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA
4801  TTACCGCCTT TGAGTGAGCT GATACCGCTC GCCGCAGCCG AACGACCGAG CGCAGCGAGT
4861  CAGTGAGCGA GGAAGCGGAA G
```

Figure 5

pMSVLSB-2: 3413 bp;
Composition 777 A; 950 C; 884 G; 802 T; 0 OTHER
Percentage: 23% A; 28% C; 26% G; 23% T; 0%OTHER

Molecular Weight (kDa): ssDNA: 1052.40 dsDNA: 2104.2
ORIGIN

```
1     AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC
61    ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC
121   TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
181   TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT
241   TTAGGTGACA CTATAGAATA CTCAAGCTAT GCATCAAGCT TGGGCCCGGT AGGGACCGAG
301   CGCTTTGATT TAAAGCCTGG TTCTGCTTTG TATGATTTAT CTAAAGCAGC CCAATCTAAA
361   GAAACCGGTC CCGGGCACTA TAAATTGCCT AACAAGTGCG ATTCATTCAT GGATCCTTTA
421   AACTCGAGTC TAGAGGGCCC GAATTCTGCA GATATCCATC ACACTGGCGG CCGCTCGAGC
481   ATGCATCTAG AGGGCCCAAT TCGCCCTATA GTGAGTCGTA TTACAATTCA CTGGCCGTCG
541   TTTTACAACG TCGTGACTGG GAAAACCCTG GCGTTACCCA ACTTAATCGC CTTGCAGCAC
601   ATCCCCCTTT CGCCAGCTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC
661   AGTTGCGCAG CCTATACGTA CGGCAGTTTA AGGTTTACAC CTATAAAAGA GAGAGCCGTT
721   ATCGTCTGTT TGTGGATGTA CAGAGTGATA TTATTGACAC GCCGGGGCGA CGGATGGTGA
781   TCCCCCTGGC CAGTGCACGT CTGCTGTCAG ATAAAGTCTC CCGTGAACTT TACCCGGTGG
841   TGCATATCGG GGATGAAAGC TGGCGCATGA TGACCACCGA TATGGCCAGT GTGCCGGTCT
901   CCGTTATCGG GGAAGAAGTG GCTGATCTCA GCCACCGCGA AAATGACATC AAAAACGCCA
961   TTAACCTGAT GTTCTGGGGA ATATAAATGT CAGGCCTGAA TGGCGAATGG ACGCGCCCTG
1021  TAGCGGCGCA TTAAGCGCGC GGGTGTGGTG GTTACGCGCA GCGTGACCGC TACACTTGCC
1081  AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGGC
1141  TTTCCCCGTC AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT TCCGATTTAG AGCTTTACGG
1201  CACCTCGACC GCAAAAAACT TGATTTGGGT GATGGTTCAC GTAGTGGGCC ATCGCCCTGA
1261  TAGACGGTTT TTCGCCCTTT GACGTTGGAG TCCACGTTCT TTAATAGTGG ACTCTTGTTC
1321  CAAACTGGAA CAACACTCAA CCCTATCGCG GTCTATTCTT TTGATTTATA AGGGATGTTG
1381  CCGATTTCGG CCTATTGGTT AAAAAATGAG CTGATTTAAC AAAAATTTTA ACAAAATTCA
1441  GAAGAACTCG TCAAGAAGGC GATAGAAGGC GATGCGCTGC GAATCGGGAG CGGCGATACC
1501  GTAAAGCACG AGGAAGCGGT CAGCCCATTC GCCGCCAAGC TCTTCAGCAA TATCACGGGT
1561  AGCCAACGCT ATGTCCTGAT AGCGGTCCGC CACACCCAGC CGGCCACAGT CGATGAATCC
1621  AGAAAAGCGG CCATTTTCCA CCATGATATT CGGCAAGCAG GCATCGCCAT GGGTCACGAC
1681  GAGATCCTCG CCGTCGGGCA TGCTCGCCTT GAGCCTGGCG AACAGTTCGG CTGGCGCGAG
1741  CCCCTGATGC TCTTCGTCCA GATCATCCTG ATCGACAAGA CCGGCTTCCA TCCGAGTACG
1801  TGCTCGCTCG ATGCGATGTT TCGCTTGGTG GTCGAATGGG CAGGTAGCCG GATCAAGCGT
1861  ATGCAGCCGC CGCATTGCAT CAGCCATGAT GGATACTTTC TCGGCAGGAG CAAGGTGAGA
1921  TGACAGGAGA TCCTGCCCCG GCACTTCGCC CAATAGCAGC CAGTCCCTTC CCGCTTCAGT
1981  GACAACGTCG AGCACAGCTG CGCAAGGAAC GCCCGTCGTG GCCAGCCACG ATAGCCGCGC
2041  TGCCTCGTCT TGCAGTTCAT TCAGGGCACC GGACAGGTCG GTCTTGACAA AAAGAACCGG
2101  GCGCCCCTGC GCTGACAGCC GGAACACGGC GGCATCAGAG CAGCCGATTG TCTGTTGTGC
2161  CCAGTCATAG CCGAATAGCC TCTCCACCCA AGCGGCCGGA GAACCTGCGT GCAATCCATC
2221  TTGTTCAATC ATGCGAAACG ATCCTCATCC TGTCTCTTGA TCAGATCTTG ATCCCCTGCG
2281  CCATCAGATC CTTGGCGGCG AGAAAGCCAT CCAGTTTACT TTGCAGGGCT TCCCAACCTT
2341  ACCAGAGGGC GCCCCAGCTG GCAATTCCGG TTCGCTTGCT GTCCATAAAA CCGCCCAGTC
2401  TAGCTATCGC CATGTAAGCC CACTGCAAGC TACCTGCTTT CTCTTTGCGC TTGCGTTTTC
2461  CCTTGTCCAG ATAGCCCAGT AGCTGACATT CATCCGGGGT CAGCACCGTT TCTGCGGACT
2521  GGCTTTCTAC GTGAAAAGGA TCTAGGTGAA GATCCTTTTT GATAATCTCA TGACCAAAAT
2581  CCCTTAACGT GAGTTTTCGT TCCACTGAGC GTCAGACCCC GTAGAAAAGA TCAAAGGATC
2641  TTCTTGAGAT CCTTTTTTTC TGCGCGTAAT CTGCTGCTTG CAAACAAAAA AACCACCGCT
2701  ACCAGCGGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT CTTTTTCCGA AGGTAACTGG
2761  CTTCAGCAGA GCGCAGATAC CAAATACTGT CCTTCTAGTG TAGCCGTAGT TAGGCCACCA
2821  CTTCAAGAAC TCTGTAGCAC CGCCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC
2881  TGCTGCCAGT GGCGATAAGT CGTGTCTTAC CGGGTTGGAC TCAAGACGAT AGTTACCGGA
```

Figure 5 (cont'd)

```
2941   TAAGGCGCAG CGGTCGGGCT GAACGGGGGG TTCGTGCACA CAGCCCAGCT TGGAGCGAAC
3001   GACCTACACC GAACTGAGAT ACCTACAGCG TGAGCTATGA GAAAGCGCCA CGCTTCCCGA
3061   AGGGAGAAAG GCGGACAGGT ATCCGGTAAG CGGCAGGGTC GGAACAGGAG AGCGCACGAG
3121   GGAGCTTCCA GGGGGAAACG CCTGGTATCT TTATAGTCCT GTCGGGTTTC GCCACCTCTG
3181   ACTTGAGCGT CGATTTTTGT GATGCTCGTC AGGGGGGCGG AGCCTATGGA AAAACGCCAG
3241   CAACGCGGCC TTTTTACGGT TCCTGGGCTT TTGCTGGCCT TTTGCTCACA TGTTCTTTCC
3301   TGCGTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC TTTGAGTGAG CTGATACCGC
3361   TCGCCGCAGC CGAACGACCG AGCGCAGCGA GTCAGTGAGC GAGGAAGCGG AAG
```

Figure 6

pMSVLSB-3:

pMSVLSB2 Apa fragment inserted: 4961 bp;
Composition 1190 A; 1276 C; 1262 G; 1233 T; 0 OTHER
Percentage: 24% A; 26% C; 25% G; 25% T; 0%OTHER

Molecular Weight (kDa): ssDNA: 1531.26 dsDNA: 3058.5
ORIGIN

```
1     AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC
61    ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC
121   TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
181   TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT
241   TTAGGTGACA CTATAGAATA CTCAAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA
301   CTAGTAACGG CCGCCAGTGT GCTGGAATTC ATGGGCAGAC CCGTCTGTAC TTTAAGAGTG
361   TTGGCAACCA GTAATGAATA AAAACTCCCG TTTTATTATA TTTGATGAAT GCTGAAAGCT
421   TACATTAATA TGTCGTGCGA TGGCACGAAA AAACACACGC AAACAATACA GGGGGGTAGT
481   CGGCGGGCGG CTAAGGGTGG TGCTCGGCGG GCAGAACATC GAAAAATCAA GATCTATATG
541   AATTACACTT CCTCCGTAGG AGGAAGCACA GGGGGAGAAT ACCACTTCTC CCCCGGCGAC
601   ATAATGTAAA TGACGCAGTT TGCCTCGAAA TACTCCAGCT GCCCTGGAGT CATTTCCTTC
661   ATCCAATCTT CATCCGAGTT GGCGAGGATT ATTGTAGGCT TAGACTTCTT CTGCACCTTT
721   TTCTTCTTAC CATACTTGGG GTTTACAATG AAATCCCTCT GACAGCCAAC TAACTGTTTC
781   CAACAAGGAC AGAATTTAAA CGGAATATCA TCTACGATGT TGTAGATTGC GTCTTCGTTG
841   TATGAAGACC AATCAACATT ATTTTGCCAG TAATTATGAA CCCCTAGGCT TCTGGCCCAA
901   GTAGATTTTC CGGTTCTTGT TGGGCCGACG ATGTAGAGGC TCTGCTTTCT TGATCTTTCA
961   TCTGATGACT GGATACAGAA TCCATCCATT GGAGGTCAGA AATTGCATCC TCGAGGGTAT
1021  AACAGGTAGG TTGAAGGAGC ATGTAAGCTT CGGGACTAAC CTGGAAGATG TTAGGCTGGA
1081  GCCAATCGTT GATTGACTCA TTACAAAGTA AATCAGGTGA GGAGGGTGGA TGAGGATTGG
1141  TGAACTCTTC CTGAATCTCA GGAAAAAGCT TATTTGCAGA GTATTCAAAA TACTGCAATT
1201  TTGTGGACCA ATCAAAGGGG AGCTCTTTCT GGATCATGGA GAGGTACTCT TCTTTGGAGG
1261  TAGCGTGTGA AATAATGTCT CGCATTATTT CATCTTTAGA AGGCTTTTTT TCCTTTACCT
1321  CTGAATCAGA TTTTCCTAGG AAGGGGGACT TCCTAGGAAT GAAAGTACCT CTCTCAAACA
1381  CAGCCAGAGG TTCCTTGAGA ATGTAATCCC TCACTCTGTT AACTGACTTG GCACTCTGAA
1441  TATTTGGGTG AAACCCATTT ATATCAAAGA ACCTTGAGTC AGATATCCTT ATCGGCTTCT
1501  CTGGCTGAAG CAATGCATGT AAATGCAAAC TTCCATCTTT ATGTGCCTCT CGGGCACATA
1561  GAATATATTT GGGAATCCAA CGAACGACGA GCTCCCAGAT CATCTGACAG GCGATTTCAG
1621  GATTTTCTGG ACACTTTGGA TAGGTTAGGA ACGTGTTAGC GTTCCTGTGT GAGAACTGAC
1681  GGTTGGATGA GGAGGAGGCC ATAGCCGACG ACGGAGGTTG AGGCTGAGGG ATGGCAGACT
1741  GGGAGCTCCA AACTCTATAG TATACCCGTG CGCCTTCGAA ATCCGCCGCT CCATTGTCTT
1801  ATAGTGGTTG TAAATGGGCC GGACCGGGCC GGCCCAGCAG GAAAAGAAGG CGCGCACTAA
1861  TATTACCGCG CCTTCTTTTC CTGCGAGGGC CCGGTAGGGA CCGAGCGCTT TGATTTAAAG
1921  CCTGGTTCTG CTTTGTATGA TTTATCTAAA GCAGCCCAAT CTAAAGAAAC CGGTCCCGGG
1981  CACTATAAAT TGCCTAACAA GTGCGATTCA TTCATGGATC CTTTAAACTC GAGTCTAGAG
2041  GGCCCAATTC GCCCTATAGT GAGTCGTATT ACAATTCACT GGCCGTCGTT TTACAACGTC
2101  GTGACTGGGA AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCCTTTCG
2161  CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC
2221  TATACGTACG GCAGTTTAAG GTTTACACCT ATAAAAGAGA GAGCCGTTAT CGTCTGTTTG
2281  TGGATGTACA GAGTGATATT ATTGACACGC CGGGGCGACG GATGGTGATC CCCCTGGCCA
2341  GTGCACGTCT GCTGTCAGAT AAAGTCTCCC GTGAACTTTA CCCGGTGGTG CATATCGGGG
2401  ATGAAAGCTG GCGCATGATG ACCACCGATA TGGCCAGTGT GCCGGTCTCC GTTATCGGGG
2461  AAGAAGTGGC TGATCTCAGC CACCGCGAAA ATGACATCAA AAACGCCATT AACCTGATGT
2521  TCTGGGGAAT ATAAATGTCA GGCCTGAATG GCGAATGGAC GCGCCCTGTA GCGGCGCATT
2581  AAGCGCGCGG GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG
2641  CCCGCTCCTT TCGCTTTCTT CCCTTCCTTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA
2701  GCTCTAAATC GGGGGCTCCC TTTAGGGTTC CGATTTAGAG CTTTACGGCA CCTCGACCGC
2761  AAAAAACTTG ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT
```

Figure 6 (cont'd)

```
2821  CGCCCTTTGA CGTTGGAGTC CACGTTCTTT AATAGTGGAC TCTTGTTCCA AACTGGAACA
2881  ACACTCAACC CTATCGCGGT CTATTCTTTT GATTTATAAG GGATGTTGCC GATTTCGGCC
2941  TATTGGTTAA AAAATGAGCT GATTTAACAA AAATTTTAAC AAAATTCAGA AGAACTCGTC
3001  AAGAAGGCGA TAGAAGGCGA TGCGCTGCGA ATCGGGAGCG GCGATACCGT AAAGCACGAG
3061  GAAGCGGTCA GCCCATTCGC CGCCAAGCTC TTCAGCAATA TCACGGGTAG CCAACGCTAT
3121  GTCCTGATAG CGGTCCGCCA CACCCAGCCG GCCACAGTCG ATGAATCCAG AAAAGCGGCC
3181  ATTTTCCACC ATGATATTCG GCAAGCAGGC ATCGCCATGG GTCACGACGA GATCCTCGCC
3241  GTCGGGCATG CTCGCCTTGA GCCTGGCGAA CAGTTCGGCT GGCGCGAGCC CCTGATGCTC
3301  TTCGTCCAGA TCATCCTGAT CGACAAGACC GGCTTCCATC CGAGTACGTG CTCGCTCGAT
3361  GCGATGTTTC GCTTGGTGGT CGAATGGGCA GGTAGCCGGA TCAAGCGTAT GCAGCCGCCG
3421  CATTGCATCA GCCATGATGG ATACTTTCTC GGCAGGAGCA AGGTGAGATG ACAGGAGATC
3481  CTGCCCCGGC ACTTCGCCCA ATAGCAGCCA GTCCCTTCCC GCTTCAGTGA CAACGTCGAG
3541  CACAGCTGCG CAAGGAACGC CCGTCGTGGC CAGCCACGAT AGCCGCGCTG CCTCGTCTTG
3601  CAGTTCATTC AGGGCACCGG ACAGGTCGGT CTTGACAAAA AGAACCGGGC GCCCCTGCGC
3661  TGACAGCCGG AACACGGCGG CATCAGAGCA GCCGATTGTC TGTTGTGCCC AGTCATAGCC
3721  GAATAGCCTC TCCACCCAAG CGGCCGGAGA ACCTGCGTGC AATCCATCTT GTTCAATCAT
3781  GCGAAACGAT CCTCATCCTG TCTCTTGATC AGATCTTGAT CCCCTGCGCC ATCAGATCCT
3841  TGGCGGCGAG AAAGCCATCC AGTTTACTTT GCAGGGCTTC CCAACCTTAC CAGAGGGCGC
3901  CCCAGCTGGC AATTCCGGTT CGCTTGCTGT CCATAAAACC GCCCAGTCTA GCTATCGCCA
3961  TGTAAGCCCA CTGCAAGCTA CCTGCTTTCT CTTTGCGCTT GCGTTTTCCC TTGTCCAGAT
4021  AGCCCAGTAG CTGACATTCA TCCGGGGTCA GCACCGTTTC TGCGGACTGG CTTTCTACGT
4081  GAAAAGGATC TAGGTGAAGA TCCTTTTTGA TAATCTCATG ACCAAAATCC CTTAACGTGA
4141  GTTTTCGTTC CACTGAGCGT CAGACCCCGT AGAAAAGATC AAAGGATCTT CTTGAGATCC
4201  TTTTTTTCTG CGCGTAATCT GCTGCTTGCA AACAAAAAAA CCACCGCTAC CAGCGGTGGT
4261  TTGTTTGCCG GATCAAGAGC TACCAACTCT TTTTCCGAAG GTAACTGGCT TCAGCAGAGC
4321  GCAGATACCA AATACTGTCC TTCTAGTGTA GCCGTAGTTA GGCCACCACT TCAAGAACTC
4381  TGTAGCACCG CCTACATACC TCGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG
4441  CGATAAGTCG TGTCTTACCG GGTTGGACTC AAGACGATAG TTACCGGATA AGGCGCAGCG
4501  GTCGGGCTGA ACGGGGGGTT CGTGCACACA GCCCAGCTTG GAGCGAACGA CCTACACCGA
4561  ACTGAGATAC CTACAGCGTG AGCTATGAGA AAGCGCCACG CTTCCCGAAG GGAGAAAGGC
4621  GGACAGGTAT CCGGTAAGCG GCAGGGTCGG AACAGGAGAG CGCACGAGGG AGCTTCCAGG
4681  GGGAAACGCC TGGTATCTTT ATAGTCCTGT CGGGTTTCGC CACCTCTGAC TTGAGCGTCG
4741  ATTTTTGTGA TGCTCGTCAG GGGGGCGGAG CCTATGGAAA AACGCCAGCA ACGCGGCCTT
4801  TTTACGGTTC CTGGGCTTTT GCTGGCCTTT TGCTCACATG TTCTTTCCTG CGTTATCCCC
4861  TGATTCTGTG GATAACCGTA TTACCGCCTT TGAGTGAGCT GATACCGCTC GCCGCAGCCG
4921  AACGACCGAG CGCAGCGAGT CAGTGAGCGA GGAAGCGGAA G
```

Figure 7

pMSVLSB4: 6309 bp;
Composition 1522 A; 1620 C; 1590 G; 1577 T; 0 OTHER
Percentage: 24% A; 26% C; 25% G; 25% T; 0%OTHER

Molecular Weight (kDa): ssDNA: 1947.08 dsDNA: 3889.6
ORIGIN

```
1       AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC
61      ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC
121     TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
181     TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT
241     TTAGGTGACA CTATAGAATA CTCAAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA
301     CTAGTCCCGA TCTAGTAACA TAGATGACAC CGCGCGCGAT AATTTATCCT AGTTTGCGCG
361     CTATATTTTG TTTTCTATCG CGTATTAAAT GTATAATTGC GGGACTCTAA TCATAAAAAC
421     CCATCTCATA AATAACGTCA TGCATTACAT GTTAATTATT ACATGCTTAA CGTAATTCAA
481     CAGAAATTAT ATGATAATCA TCGACAGACC GGCAACAGGA TTCAATCTTA AGAAACTTTA
541     TTGCCAAATG TTTGAACGAT CGGGGAAATT CGCTCGAGTT AATTAAGCGG CCGCCTCAAA
601     AAGGATCTTC ACCTAGATCC TTTTAAATTA AAAATGAAGT TTTAGCACGT GTCAGTCCTG
661     CTCCTCGGCC ACGAAGTGCA CGCAGTTGCC GGCCGGGTCG CGCAGGGCGA ACTCCCGCCC
721     CCACGGCTGC TCGCCGATCT CGGTCATGGC CGGCCCGGAG GCGTCCCGGA AGTTCGTGGA
781     CACGACCTCC GACCACTCGG CGTACAGCTC GTCCAGGCCG CGCACCCACA CCCAGGCCAG
841     GGTGTTGTCC GGCACCACCT GGTCCTGGAC CGCGCTGATG AACAGGGTCA CGTCGTCCCG
901     GACCACACCG GCGAAGTCGT CCTCCACGAA GTCCCGGGAG AACCCGAGCC GGTCGGTCCA
961     GAACTCGACC GCTCCGGCGA CGTCGCGCGC GGTGAGCACC GGAACGGCAC TGGTCAACTT
1021    GGCCATGGTG GCCCTCCTCA CGTGCTATTA TTGAAGCATT TATCAGGGTT ATTGTCTCAT
1081    GAGCGGATAC ATATTTGAAT GTATTTAGAA AAATAAACAA ATAGGGGTTC CGCGCACATT
1141    TCCCCGAAAA GTGCCACCTG TATGCGGTGT GAAATACCGC ACAGATGCGT AAGGAGAAAA
1201    TACCGCATCA GGCGAAATTG TAAACGCGGC CGCTTAATTA AGTCGACGTC CTCTCCAAAT
1261    GAAATGAACT TCCTTATATA GAGGAAGGGT CTTGCGAAGG ATAGTGGGAT TGTGCGTCAT
1321    CCCTTACGTC AGTGGAGATA TCACATCAAT CCACTTGCTT TGAAGACGTG GTTGGAACGT
1381    CTTCTTTTTC CACGTAGCTC CTCGTGGGTG GGGGTCCATC TTTGGGACCA CTGTCGGCAG
1441    AGGCATCTTG AACGATAGCC TTTCCTTATC GCAATGATGG CATTTGTAGG TGCCACCTTC
1501    CTTTTCTACT GTCCTTTTGA TGAAGTGACA GATAGCTGGG CAATGGAATC CGAGGAGGTT
1561    TCCCGATATT ACCCTTTGTT GAAAAGTCTC AATAGCCCTT TGGTCTTCTG AGACTGTATC
1621    TTTGATATTC TTGGAGTAGA CGAGAGAGTG TCGTGCTCCA CCATGTTGAC GAATTCATGG
1681    GCAGACCCGT CTGTACTTTA AGAGTGTTGG CAACCAGTAA TGAATAAAAA CTCCCGTTTT
1741    ATTATATTTG ATGAATGCTG AAAGCTTACA TTAATATGTC GTGCGATGGC ACGAAAAAAC
1801    ACACGCAAAC AATACAGGGG GGTAGTCGGC GGGCGGCTAA GGGTGGTGCT CGGCGGGCAG
1861    AACATCGAAA AATCAAGATC TATATGAATT ACACTTCCTC CGTAGGAGGA AGCACAGGGG
1921    GAGAATACCA CTTCTCCCCC GGCGACATAA TGTAAATGAC GCAGTTTGCC TCGAAATACT
1981    CCAGCTGCCC TGGAGTCATT TCCTTCATCC AATCTTCATC CGAGTTGGCG AGGATTATTG
2041    TAGGCTTAGA CTTCTTCTGC ACCTTTTTCT TCTTACCATA CTTGGGGTTT ACAATGAAAT
2101    CCCTCTGACA GCCAACTAAC TGTTTCCAAC AAGGACAGAA TTTAAACGGA ATATCATCTA
2161    CGATGTTGTA GATTGCGTCT TCGTTGTATG AAGACCAATC AACATTATTT TGCCAGTAAT
2221    TATGAACCCC TAGGCTTCTG GCCCAAGTAG ATTTTCCGGT TCTTGTTGGG CCGACGATGT
2281    AGAGGCTCTG CTTTCTTGAT CTTTCATCTG ATGACTGGAT ACAGAATCCA TCCATTGGAG
2341    GTCAGAAATT GCATCCTCGA GGGTATAACA GGTAGGTTGA AGGAGCATGT AAGCTTCGGG
2401    ACTAACCTGG AAGATGTTAG GCTGGAGCCA ATCGTTGATT GACTCATTAC AAAGTAAATC
2461    AGGTGAGGAG GGTGGATGAG GATTGGTGAA CTCTTCCTGA ATCTCAGGAA AAAGCTTATT
2521    TGCAGAGTAT TCAAAATACT GCAATTTTGT GGACCAATCA AAGGGGAGCT CTTTCTGGAT
2581    CATGGAGAGG TACTCTTCTT TGGAGGTAGC GTGTGAAATA ATGTCTCGCA TTATTTCATC
2641    TTTAGAAGGC TTTTTTTCCT TTACCTCTGA ATCAGATTTT CCTAGGAAGG GGGACTTCCT
2701    AGGAATGAAA GTACCTCTCT CAAACACAGC CAGAGGTTCC TTGAGAATGT AATCCCTCAC
2761    TCTGTTAACT GACTTGGCAC TCTGAATATT TGGGTGAAAC CCATTTATAT CAAAGAACCT
2821    TGAGTCAGAT ATCGTTATCG GCTTCTCTGG CTGAAGCAAT GCATGTAAAT GCAAACTTCC
2881    ATCTTTATGT GCCTCTCGGG CACATAGAAT ATATTTGGGA ATCCAACGAA CGACGAGCTC
```

Figure 7 (cont'd)

```
2941  CCAGATCATC TGACAGGCGA TTTCAGGATT TTCTGGACAC TTTGGATAGG TTAGGAACGT
3001  GTTAGCGTTC CTGTGTGAGA ACTGACGGTT GGATGAGGAG GAGGCCATAG CCGACGACGG
3061  AGGTTGAGGC TGAGGGATGG CAGACTGGGA GCTCCAAACT CTATAGTATA CCCGTGCGCC
3121  TTCGAAATCC GCCGCTCCAT TGTCTTATAG TGGTTGTAAA TGGGCCGGAC CGGGCCGGCC
3181  CAGCAGGAAA AGAAGGCGCG CACTAATATT ACCGCGCCTT CTTTTCCTGC GAGGGCCCGG
3241  GGTAGGGACC GAGCGCTTTG ATTTAAAGCC TGGTTCTGCT TTGTATGATT TATCTAAAGC
3301  AGCCCAATCT AAAGAAACCG GTCCCGGGCA CTATAAATTG CCTAACAAGT GCGATTCATT
3361  CATGGATCCT TTAAACTCGA GTCTAGAGGG CCCAATTCGC CCTATAGTGA GTCGTATTAC
3421  AATTCACTGG CCGTCGTTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT TACCCAACTT
3481  AATCGCCTTG CAGCACATCC CCCTTTCGCC AGCTGGCGTA ATAGCGAAGA GGCCCGCACC
3541  GATCGCCCTT CCCAACAGTT GCGCAGCCTA TACGTACGGC AGTTTAAGGT TTACACCTAT
3601  AAAAGAGAGA GCCGTTATCG TCTGTTTGTG GATGTACAGA GTGATATTAT TGACACGCCG
3661  GGGCGACGGA TGGTGATCCC CCTGGCCAGT GCACGTCTGC TGTCAGATAA AGTCTCCCGT
3721  GAACTTTACC CGGTGGTGCA TATCGGGGAT GAAAGCTGGC GCATGATGAC CACCGATATG
3781  GCCAGTGTGC CGGTCTCCGT TATCGGGGAA GAAGTGGCTG ATCTCAGCCA CCGCGAAAAT
3841  GACATCAAAA ACGCCATTAA CCTGATGTTC TGGGGAATAT AAATGTCAGG CCTGAATGGC
3901  GAATGGACGC GCCCTGTAGC GGCGCATTAA GCGCGCGGGT GTGGTGGTTA CGCGCAGCGT
3961  GACCGCTACA CTTGCCAGCG CCCTAGCGCC CGCTCCTTTC GCTTTCTTCC CTTCCTTTCT
4021  CGCCACGTTC GCCGGCTTTC CCCGTCAAGC TCTAAATCGG GGGCTCCCTT TAGGGTTCCG
4081  ATTTAGAGCT TTACGGCACC TCGACCGCAA AAAACTTGAT TTGGGTGATG GTTCACGTAG
4141  TGGGCCATCG CCCTGATAGA CGGTTTTTCG CCCTTTGACG TTGGAGTCCA CGTTCTTTAA
4201  TAGTGGACTC TTGTTCCAAA CTGGAACAAC ACTCAACCCT ATCGCGGTCT ATTCTTTTGA
4261  TTTATAAGGG ATGTTGCCGA TTTCGGCCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA
4321  ATTTTAACAA AATTCAGAAG AACTCGTCAA GAAGGCGATA GAAGGCGATG CGCTGCGAAT
4381  CGGGAGCGGC GATACCGTAA AGCACGAGGA AGCGGTCAGC CCATTCGCCG CCAAGCTCTT
4441  CAGCAATATC ACGGGTAGCC AACGCTATGT CCTGATAGCG GTCCGCCACA CCCAGCCGGC
4501  CACAGTCGAT GAATCCAGAA AAGCGGCCAT TTTCCACCAT GATATTCGGC AAGCAGGCAT
4561  CGCCATGGGT CACGACGAGA TCCTCGCCGT CGGGCATGCT CGCCTTGAGC CTGGCGAACA
4621  GTTCGGCTGG CGCGAGCCCC TGATGCTCTT CGTCCAGATC ATCCTGATCG ACAAGACCGG
4681  CTTCCATCCG AGTACGTGCT CGCTCGATGC GATGTTTCGC TTGGTGGTCG AATGGGCAGG
4741  TAGCCGGATC AAGCGTATGC AGCCGCCGCA TTGCATCAGC CATGATGGAT ACTTTCTCGG
4801  CAGGAGCAAG GTGAGATGAC AGGAGATCCT GCCCCGGCAC TTCGCCCAAT AGCAGCCAGT
4861  CCCTTCCCGC TTCAGTGACA ACGTCGAGCA CAGCTGCGCA AGGAACGCCC GTCGTGGCCA
4921  GCCACGATAG CCGCGCTGCC TCGTCTTGCA GTTCATTCAG GGCACCGGAC AGGTCGGTCT
4981  TGACAAAAAG AACCGGGCGC CCCTGCGCTG ACAGCCGGAA CACGGCGGCA TCAGAGCAGC
5041  CGATTGTCTG TTGTGCCCAG TCATAGCCGA ATAGCCTCTC CACCCAAGCG GCCGGAGAAC
5101  CTGCGTGCAA TCCATCTTGT TCAATCATGC GAAACGATCC TCATCCTGTC TCTTGATCAG
5161  ATCTTGATCC CCTGCGCCAT CAGATCCTTG GCGGCGAGAA AGCCATCCAG TTTACTTTGC
5221  AGGGCTTCCC AACCTTACCA GAGGGCGCCC CAGCTGGCAA TTCCGGTTCG CTTGCTGTCC
5281  ATAAAACCGC CCAGTCTAGC TATCGCCATG TAAGCCCACT GCAAGCTACC TGCTTTCTCT
5341  TTGCGCTTGC GTTTTCCCTT GTCCAGATAG CCCAGTAGCT GACATTCATC CGGGGTCAGC
5401  ACCGTTTCTG CGGACTGGCT TTCTACGTGA AAAGGATCTA GGTGAAGATC CTTTTTGATA
5461  ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG
5521  AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA
5581  CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT
5641  TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA TACTGTCCTT CTAGTGTAGC
5701  CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
5761  TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA
5821  GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC
5881  CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA
5941  GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA
6001  CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG
6061  GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC
6121  TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGGCTTTTGC TGGCCTTTTG
6181  CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG
6241  AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG
6301  AAGCGGAAG
```

Figure 8

pMSVLSB-5: 8043 bp;
Composition 1983 A; 1992 C; 2011 G; 2057 T; 0 OTHER
Percentage: 25% A; 25% C; 25% G; 26% T; 0%OTHER

Molecular Weight (kDa): ssDNA: 2483.31 dsDNA: 4958.5
ORIGIN

```
1     AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC
61    ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC
121   TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
181   TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT
241   TTAGGTGACA CTATAGAATA CTCAAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA
301   CTAGTAACGG CCGCCAGTGT GCTGGAATTC ATGGGCAGAC CCGTCTGTAC TTTAAGAGTG
361   TTGGCAACCA GTAATGAATA AAAACTCCCG TTTTATTATA TTTGATGAAT GCTGAAAGCT
421   TACATTAATA TGTCGTGCGA TGGCACGAAA AAACACACGC AAACAATACA GGGGGGTAGT
481   CGGCGGGCGG CTAAGGGTGG TGCTCGGCGG GCAGAACATC GAAAAATCAA GATCTATATG
541   AATTACACTT CCTCCGTAGG AGGAAGCACA GGGGGAGAAT ACCACTTCTC CCCCGGCGAC
601   ATAATGTAAA TGACGCAGTT TGCCTCGAAA TACTCCAGCT GCCCTGGAGT CATTTCCTTC
661   ATCCAATCTT CATCCGAGTT GGCGAGGATT ATTGTAGGCT TAGACTTCTT CTGCACCTTT
721   TTCTTCTTAC CATACTTGGG GTTTACAATG AAATCCCTCT GACAGCCAAC TAACTGTTTC
781   CAACAAGGAC AGAATTTAAA CGGAATATCA TCTACGATGT TGTAGATTGC GTCTTCGTTG
841   TATGAAGACC AATCAACATT ATTTTGCCAG TAATTATGAA CCCCTAGGCT TCTGGCCCAA
901   GTAGATTTTC CGGTTCTTGT TGGGCCGACG ATGTAGAGGC TCTGCTTTCT TGATCTTTCA
961   TCTGATGACT GGATACAGAA TCCATCCATT GGAGGTCAGA AATTGCATCC TCGAGGGTAT
1021  AACAGGTAGG TTGAAGGAGC ATGTAAGCTT CGGGACTAAC CTGGAAGATG TTAGGCTGGA
1081  GCCAATCGTT GATTGACTCA TTACAAAGTA AATCAGGTGA GGAGGGTGGA TGAGGATTGG
1141  TGAACTCTTC CTGAATCTCA GGAAAAAGCT TATTTGCAGA GTATTCAAAA TACTGCAATT
1201  TTGTGGACCA ATCAAAGGGG AGCTCTTTCT GGATCATGGA GAGGTACTCT TCTTTGGAGG
1261  TAGCGTGTGA AATAATGTCT CGCATTATTT CATCTTTAGA AGGCTTTTTT TCCTTTACCT
1321  CTGAATCAGA TTTTCCTAGG AAGGGGGACT TCCTAGGAAT GAAAGTACCT CTCTCAAACA
1381  CAGCCAGAGG TTCCTTGAGA ATGTAATCCC TCACTCTGTT AACTGACTTG GCACTCTGAA
1441  TATTTGGGTG AAACCCATTT ATATCAAAGA ACCTTGAGTC AGATATCCTT ATCGGCTTCT
1501  CTGGCTGAAG CAATGCATGT AAATGCAAAC TTCCATCTTT ATGTGCCTCT CGGGCACATA
1561  GAATATATTT GGGAATCCAA CGAACGACGA GCTCCCAGAT CATCTGACAG GCGATTTCAG
1621  GATTTTCTGG ACACTTTGGA TAGGTTAGGA ACGTGTTAGC GTTCCTGTGT GAGAACTGAC
1681  GGTTGGATGA GGAGGAGGCC ATAGCCGACG ACGGAGGTTG AGGCTGAGGG ATGGCAGACT
1741  GGGAGCTCCA AACTCTATAG TATACCCGTG CGCCTTCGAA ATCCGCCGCT CCATTGTCTT
1801  ATAGTGGTTG TAAATGGGCC GGACCGGGCC GGCCCAGCAG GAAAAGAAGG CGCGCACTAA
1861  TATTACCGCG CCTTCTTTTC CTGCGAGGGC CCGGTAGGGA CCGAGCGCTT TGATTTAAAG
1921  CCTGGTTCTG CTTTGTATGA TTTATCTAAA GCAGCCCAAT CTAAAGAAAC CGGTCCCGGG
1981  CACTATAAAT TGCCTAACAA GTGCGATTCA TTCATGGATC CTTTAAACTC GAGTCTAGTC
2041  CCGATCTAGT AACATAGATG ACACCGCGCG CGATAATTTA TCCTAGTTTG CGCGCTATAT
2101  TTTGTTTTCT ATCGCGTATT AAATGTATAA TTGCGGGACT CTAATCATAA AAACCCATCT
2161  CATAAATAAC GTCATGCATT ACATGTTAAT TATTACATGC TTAACGTAAT TCAACAGAAA
2221  TTATATGATA ATCATCGACA GACCGGCAAC AGGATTCAAT CTTAAGAAAC TTTATTGCCA
2281  AATGTTTGAA CGATCGGGGA AATTCGCTCG AGTTAATTAA GCGGCCGCCT CAAAAAGGAT
2341  CTTCACCTAG ATCCTTTTAA ATTAAAAATG AAGTTTTAGC ACGTGTCAGT CCTGCTCCTC
2401  GGCCACGAAG TGCACGCAGT TGCCGGCCGG GTCGCGCAGG GCGAACTCCC GCCCCCACGG
2461  CTGCTCGCCG ATCTCGGTCA TGGCCGGCCC GGAGGCGTCC CGGAAGTTCG TGGACACGAC
2521  CTCCGACCAC TCGGCGTACA GCTCGTCCAG GCCGCGCACC CACACCCAGG CCAGGGTGTT
2581  GTCCGGCACC ACCTGGTCCT GGACCGCGCT GATGAACAGG GTCACGTCGT CCCGGACCAC
2641  ACCGGCGAAG TCGTCCTCCA CGAAGTCCCG GGAGAACCCG AGCCGGTCGG TCCAGAACTC
2701  GACCGCTCCG GCGACGTCGC GCGCGGTGAG CACCGGAACG GCACTGGTCA ACTTGGCCAT
2761  GGTGGCCCTC CTCACGTGCT ATTATTGAAG CATTTATCAG GGTTATTGTC TCATGAGCGG
2821  ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG GTTCCGCGCA CATTTCCCCG
2881  AAAAGTGCCA CCTGTATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC
```

Figure 8 (cont'd)

```
2941  ATCAGGCGAA ATTGTAAACG CGGCCGCTTA ATTAAGTCGA CGTCCTCTCC AAATGAAATG
3001  AACTTCCTTA TATAGAGGAA GGGTCTTGCG AAGGATAGTG GGATTGTGCG TCATCCCTTA
3061  CGTCAGTGGA GATATCACAT CAATCCACTT GCTTTGAAGA CGTGGTTGGA ACGTCTTCTT
3121  TTTCCACGTA GCTCCTCGTG GGTGGGGGTC CATCTTTGGG ACCACTGTCG GCAGAGGCAT
3181  CTTGAACGAT AGCCTTTCCT TATCGCAATG ATGGCATTTG TAGGTGCCAC CTTCCTTTTC
3241  TACTGTCCTT TTGATGAAGT GACAGATAGC TGGGCAATGG AATCCGAGGA GGTTTCCCGA
3301  TATTACCCTT TGTTGAAAAG TCTCAATAGC GCTTTGGTCT TCTGAGACTG TATCTTTGAT
3361  ATTCTTGGAG TAGACGAGAG AGTGTCGTGC TCCACCATGT TGACGAATTC ATGGGCAGAC
3421  CCGTCTGTAC TTTAAGAGTG TTGGCAACCA GTAATGAATA AAAACTCCCG TTTTATTATA
3481  TTTGATGAAT GCTGAAAGCT TACATTAATA TGTCGTGCGA TGGCACGAAA AAACACACGC
3541  AAACAATACA GGGGGGTAGT CGGCGGGCGG CTAAGGGTGG TGCTCGGCGG GCAGAACATC
3601  GAAAAATCAA GATCTATATG AATTACACTT CCTCCGTAGG AGGAAGCACA GGGGGAGAAT
3661  ACCACTTCTC CCCCGGCGAC ATAATGTAAA TGACGCAGTT TGCCTCGAAA TACTCCAGCT
3721  GCCCTGGAGT CATTTCCTTC ATCCAATCTT CATCCGAGTT GGCGAGGATT ATTGTAGGCT
3781  TAGACTTCTT CTGCACCTTT TTCTTCTTAC CATACTTGGG GTTTACAATG AAATCCCTCT
3841  GACAGCCAAC TAACTGTTTC CAACAAGGAC AGAATTTAAA CGGAATATCA TCTACGATGT
3901  TGTAGATTGC GTCTTCGTTG TATGAAGACC AATCAACATT ATTTTGCCAG TAATTATGAA
3961  CCCCTAGGCT TCTGGCCCAA GTAGATTTTC CGGTTCTTGT TGGGCCGACG ATGTAGAGGC
4021  TCTGCTTTCT TGATCTTTCA TCTGATGACT GGATACAGAA TCCATCCATT GGAGGTCAGA
4081  AATTGCATCC TCGAGGGTAT AACAGGTAGG TTGAAGGAGC ATGTAAGCTT CGGGACTAAC
4141  CTGGAAGATG TTAGGCTGGA GCCAATCGTT GATTGACTCA TTACAAAGTA AATCAGGTGA
4201  GGAGGGTGGA TGAGGATTGG TGAACTCTTC CTGAATCTCA GGAAAAAGCT TATTTGCAGA
4261  GTATTCAAAA TACTGCAATT TTGTGGACCA ATCAAAGGGG AGCTCTTTCT GGATCATGGA
4321  GAGGTACTCT TCTTTGGAGG TAGCGTGTGA AATAATGTCT CGCATTATTT CATCTTTAGA
4381  AGGCTTTTTT TCCTTTACCT CTGAATCAGA TTTTCCTAGG AAGGGGGACT TCCTAGGAAT
4441  GAAAGTACCT CTCTCAAACA CAGCCAGAGG TTCCTTGAGA ATGTAATCCC TCACTCTGTT
4501  AACTGACTTG GCACTCTGAA TATTTGGGTG AAACCCATTT ATATCAAAGA ACCTTGAGTC
4561  AGATATCCTT ATCGGCTTCT CTGGCTGAAG CAATGCATGT AAATGCAAAC TTCCATCTTT
4621  ATGTGCCTCT CGGGCACATA GAATATATTT GGGAATCCAA CGAACGACGA GCTCCCAGAT
4681  CATCTGACAG GCGATTTCAG GATTTTCTGG ACACTTTGGA TAGGTTAGGA ACGTGTTAGC
4741  GTTCCTGTGT GAGAACTGAC GGTTGGATGA GGAGGAGGCC ATAGCCGACG ACGGAGGTTG
4801  AGGCTGAGGG ATGGCAGACT GGGAGCTCCA AACTCTATAG TATACCCGTG CGCCTTCGAA
4861  ATCCGCCGCT CCATTGTCTT ATAGTGGTTG TAAATGGGCC GGACCGGGCC GGCCCAGCAG
4921  GAAAAGAAGG CGCGCACTAA TATTACCGCG CCTTCTTTTC CTGCGAGGGC CCGGGGTAGG
4981  GACCGAGCGC TTTGATTTAA AGCCTGGTTC TGCTTTGTAT GATTTATCTA AAGCAGCCCA
5041  ATCTAAAGAA ACCGGTCCCG GGCACTATAA ATTGCCTAAC AAGTGCGATT CATTCATGGA
5101  TCCTTTAAAC TCGAGTCTAG AGGGCCCAAT TCGCCCTATA GTGAGTCGTA TTACAATTCA
5161  CTGGCCGTCG TTTTACAACG TCGTGACTGG GAAAACCCTG GCGTTACCCA ACTTAATCGC
5221  CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC
5281  CCTTCCCAAC AGTTGCGCAG CCTATACGTA CGGCAGTTTA AGGTTTACAC CTATAAAAGA
5341  GAGAGCCGTT ATCGTCTGTT TGTGGATGTA CAGAGTGATA TTATTGACAC GCCGGGGCGA
5401  CGGATGGTGA TCCCCCTGGC CAGTGCACGT CTGCTGTCAG ATAAAGTCTC CCGTGAACTT
5461  TACCCGGTGG TGCATATCGG GGATGAAAGC TGGCGCATGA TGACCACCGA TATGGCCAGT
5521  GTGCCGGTCT CCGTTATCGG GGAAGAAGTG GCTGATCTCA GCCACCGCGA AAATGACATC
5581  AAAAACGCCA TTAACCTGAT GTTCTGGGGA ATATAAATGT CAGGCCTGAA TGGCGAATGG
5641  ACGCGCCCTG TAGCGGCGCA TTAAGCGCGC GGGTGTGGTG GTTACGCGCA GCGTGACCGC
5701  TACACTTGCC AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC
5761  GTTCGCCGGC TTTCCCCGTC AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT TCCGATTTAG
5821  AGCTTTACGG CACCTCGACC GCAAAAAACT TGATTTGGGT GATGGTTCAC GTAGTGGGCC
5881  ATCGCCCTGA TAGACGGTTT TTCGCCCTTT GACGTTGGAG TCCACGTTCT TTAATAGTGG
5941  ACTCTTGTTC CAAACTGGAA CAACACTCAA CCCTATCGCG GTCTATTCTT TTGATTTATA
6001  AGGGATGTTG CCGATTTCGG CCTATTGGTT AAAAAATGAG CTGATTTAAC AAAAATTTTA
6061  ACAAAATTCA GAAGAACTCG TCAAGAAGGC GATAGAAGGC GATGCGCTGC GAATCGGGAG
6121  CGGCGATACC GTAAAGCACG AGGAAGCGGT CAGCCCATTC GCCGCCAAGC TCTTCAGCAA
6181  TATCACGGGT AGCCAACGCT ATGTCCTGAT AGCGGTCCGC CACACCCAGC CGGCCACAGT
6241  CGATGAATCC AGAAAAGCGG CCATTTTCCA CCATGATATT CGGCAAGCAG GCATCGCCAT
6301  GGGTCACGAC GAGATCCTCG CCGTCGGGCA TGCTCGCCTT GAGCCTGGCG AACAGTTCGG
```

Figure 8 (cont'd)

```
6361  CTGGCGCGAG CCCCTGATGC TCTTCGTCCA GATCATCCTG ATCGACAAGA CCGGCTTCCA
6421  TCCGAGTACG TGCTCGCTCG ATGCGATGTT TCGCTTGGTG GTCGAATGGG CAGGTAGCCG
6481  GATCAAGCGT ATGCAGCCGC CGCATTGCAT CAGCCATGAT GGATACTTTC TCGGCAGGAG
6541  CAAGGTGAGA TGACAGGAGA TCCTGCCCCG GCACTTCGCC CAATAGCAGC CAGTCCCTTC
6601  CCGCTTCAGT GACAACGTCG AGCACAGCTG CGCAAGGAAC GCCCGTCGTG GCCAGCCACG
6661  ATAGCCGCGC TGCCTCGTCT TGCAGTTCAT TCAGGGCACC GGACAGGTCG GTCTTGACAA
6721  AAAGAACCGG GCGCCCCTGC GCTGACAGCC GGAACACGGC GGCATCAGAG CAGCCGATTG
6781  TCTGTTGTGC CCAGTCATAG CCGAATAGCC TCTCCACCCA AGCGGCCGGA GAACCTGCGT
6841  GCAATCCATC TTGTTCAATC ATGCGAAACG ATCCTCATCC TGTCTCTTGA TCAGATCTTG
6901  ATCCCCTGCG CCATCAGATC CTTGGCGGCG AGAAAGCCAT CCAGTTTACT TTGCAGGGCT
6961  TCCCAACCTT ACCAGAGGGC GCCCCAGCTG GCAATTCCGG TTCGCTTGCT GTCCATAAAA
7021  CCGCCCAGTC TAGCTATCGC CATGTAAGCC CACTGCAAGC TACCTGCTTT CTCTTTGCGC
7081  TTGCGTTTTC CCTTGTCCAG ATAGCCCAGT AGCTGACATT CATCCGGGGT CAGCACCGTT
7141  TCTGCGGACT GGCTTTCTAC GTGAAAAGGA TCTAGGTGAA GATCCTTTTT GATAATCTCA
7201  TGACCAAAAT CCCTTAACGT GAGTTTTCGT TCCACTGAGC GTCAGACCCC GTAGAAAAGA
7261  TCAAAGGATC TTCTTGAGAT CCTTTTTTTC TGCGCGTAAT CTGCTGCTTG CAAACAAAAA
7321  AACCACCGCT ACCAGCGGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT CTTTTTCCGA
7381  AGGTAACTGG CTTCAGCAGA GCGCAGATAC CAAATACTGT CCTTCTAGTG TAGCCGTAGT
7441  TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCCTACATA CCTCGCTCTG CTAATCCTGT
7501  TACCAGTGGC TGCTGCCAGT GGCGATAAGT CGTGTCTTAC CGGGTTGGAC TCAAGACGAT
7561  AGTTACCGGA TAAGGCGCAG CGGTCGGGCT GAACGGGGGG TTCGTGCACA CAGCCCAGCT
7621  TGGAGCGAAC GACCTACACC GAACTGAGAT ACCTACAGCG TGAGCTATGA GAAAGCGCCA
7681  CGCTTCCCGA AGGGAGAAAG GCGGACAGGT ATCCGGTAAG CGGCAGGGTC GGAACAGGAG
7741  AGCGCACGAG GGAGCTTCCA GGGGGAAACG CCTGGTATCT TTATAGTCCT GTCGGGTTTC
7801  GCCACCTCTG ACTTGAGCGT CGATTTTTGT GATGCTCGTC AGGGGGGCGG AGCCTATGGA
7861  AAAACGCCAG CAACGCGGCC TTTTTACGGT TCCTGGGCTT TTGCTGGCCT TTTGCTCACA
7921  TGTTCTTTCC TGCGTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC TTTGAGTGAG
7981  CTGATACCGC TCGCCGCAGC CGAACGACCG AGCGCAGCGA GTCAGTGAGC GAGGAAGCGG
8041  AAG
```

Figure 9

pMSVLSB-6: 7404 bp;

Composition 1839 A; 1794 C; 1835 G; 1936 T; 0 OTHER
Percentage: 25% A; 24% C; 25% G; 26% T; 0%OTHER

Molecular Weight (kDa): ssDNA: 2286.33 dsDNA: 4564.5

ORIGIN

```
1     AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC
61    ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC
121   TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
181   TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT
241   TTAGGTGACA CTATAGAATA CTCAAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA
301   CTAGTAACGG CCGCCAGTGT GCTGGAATTC ATGGGCAGAC CCGTCTGTAC TTTAAGAGTG
361   TTGGCAACCA GTAATGAATA AAAACTCCCG TTTTATTATA TTTGATGAAT GCTGAAAGCT
421   TACATTAATA TGTCGTGCGA TGGCACGAAA AAACACACGC AAACAATACA GGGGGGTAGT
481   CGGCGGGCGG CTAAGGGTGG TGCTCGGCGG GCAGAACATC GAAAAATCAA GATCTATATG
541   AATTACACTT CCTCCGTAGG AGGAAGCACA GGGGGAGAAT ACCACTTCTC CCCCGGCGAC
601   ATAATGTAAA TGACGCAGTT TGCCTCGAAA TACTCCAGCT GCCCTGGAGT CATTTCCTTC
661   ATCCAATCTT CATCCGAGTT GGCGAGGATT ATTGTAGGCT TAGACTTCTT CTGCACCTTT
721   TTCTTCTTAC CATACTTGGG GTTTACAATG AAATCCCTCT GACAGCCAAC TAACTGTTTC
781   CAACAAGGAC AGAATTTAAA CGGAATATCA TCTACGATGT TGTAGATTGC GTCTTCGTTG
841   TATGAAGACC AATCAACATT ATTTTGCCAG TAATTATGAA CCCCTAGGCT TCTGGCCCAA
901   GTAGATTTTC CGGTTCTTGT TGGGCCGACG ATGTAGAGGC TCTGCTTTCT TGATCTTTCA
961   TCTGATGACT GGATACAGAA TCCATCCATT GGAGGTCAGA AATTGCATCC TCGAGGGTAT
1021  AACAGGTAGG TTGAAGGAGC ATGTAAGCTT CGGGACTAAC CTGGAAGATG TTAGGCTGGA
1081  GCCAATCGTT GATTGACTCA TTACAAAGTA AATCAGGTGA GGAGGGTGGA TGAGGATTGG
1141  TGAACTCTTC CTGAATCTCA GGAAAAAGCT TATTTGCAGA GTATTCAAAA TACTGCAATT
1201  TTGTGGACCA ATCAAAGGGG AGCTCTTTCT GGATCATGGA GAGGTACTCT TCTTTGGAGG
1261  TAGCGTGTGA AATAATGTCT CGCATTATTT CATCTTTAGA AGGCTTTTTT TCCTTTACCT
1321  CTGAATCAGA TTTTCCTAGG AAGGGGGACT TCCTAGGAAT GAAAGTACCT CTCTCAAACA
1381  CAGCCAGAGG TTCCTTGAGA ATGTAATCCC TCACTCTGTT AACTGACTTG GCACTCTGAA
1441  TATTTGGGTG AAACCCATTT ATATCAAAGA ACCTTGAGTC AGATATCCTT ATCGGCTTCT
1501  CTGGCTGAAG CAATGCATGT AAATGCAAAC TTCCATCTTT ATGTGCCTCT CGGGCACATA
1561  GAATATATTT GGGAATCCAA CGAACGACGA GCTCCCAGAT CATCTGACAG GCGATTTCAG
1621  GATTTTCTGG ACACTTTGGA TAGGTTAGGA ACGTGTTAGC GTTCCTGTGT GAGAACTGAC
1681  GGTTGGATGA GGAGGAGGCC ATAGCCGACG ACGGAGGTTG AGGCTGAGGG ATGGCAGACT
1741  GGGAGCTCCA AACTCTATAG TATACCCGTG CGCCTTCGAA ATCCGCCGCT CCATTGTCTT
1801  ATAGTGGTTG TAAATGGGCC GGACCGGGCC GGCCCAGCAG GAAAAGAAGG CGCGCACTAA
1861  TATTACCGCG CCTTCTTTTC CTGCGAGGGC CCGGTAGGGA CCGAGCGCTT TGATTTAAAG
1921  CCTGGTTCTG CTTTGTATGA TTTATCTAAA GCAGCCCAAT CTAAAGAAAC CGGTCCCGGG
1981  CACTATAAAT TGCCTAACAA GTGCGATTCA TTCATGGATC CTTTAAACTC GAGTCTAGTC
2041  CCGATCTAGT AACATAGATG ACACCGCGCG CGATAATTTA TCCTAGTTTG CGCGCTATAT
2101  TTTGTTTTCT ATCGCGTATT AAATGTATAA TTGCGGGACT CTAATCATAA AAACCCATCT
2161  CATAAATAAC GTCATGCATT ACATGTTAAT TATTACATGC TTAACGTAAT TCAACAGAAA
2221  TTATATGATA ATCATCGACA GACCGGCAAC AGGATTCAAT CTTAAGAAAC TTTATTGCCA
2281  AATGTTTGAA CGATCGGGGA AATTCGCTCG AGTTAATTAA GCGGCCGCTT AATTAAGTCG
2341  ACGTCCTCTC CAAATGAAAT GAACTTCCTT ATATAGAGGA AGGGTCTTGC GAAGGATAGT
2401  GGGATTGTGC GTCATCCCTT ACGTCAGTGG AGATATCACA TCAATCCACT TGCTTTGAAG
2461  ACGTGGTTGG AACGTCTTCT TTTTCCACGT AGCTCCTCGT GGGTGGGGGT CCATCTTTGG
2521  GACCACTGTC GGCAGAGGCA TCTTGAACGA TAGCCTTTCC TTATCGCAAT GATGGCATTT
2581  GTAGGTGCCA CCTTCCTTTT CTACTGTCCT TTTGATGAAG TGACAGATAG CTGGGCAATG
2641  GAATCCGAGG AGGTTTCCCG ATATTACCCT TTGTTGAAAA GTCTCAATAG CCCTTTGGTC
2701  TTCTGAGACT GTATCTTTGA TATTCTTGGA GTAGACGAGA GAGTGTCGTG CTCCACCATG
2761  TTGACGAATT CATGGGCAGA CCCGTCTGTA CTTTAAGAGT GTTGGCAACC AGTAATGAAT
2821  AAAAACTCCC GTTTTATTAT ATTTGATGAA TGCTGAAAGC TTACATTAAT ATGTCGTGCG
```

Figure 9 (cont'd)

```
2881  ATGGCACGAA AAAACACACG CAAACAATAC AGGGGGGTAG TCGGCGGGCG GCTAAGGGTG
2941  GTGCTCGGCG GGCAGAACAT CGAAAAATCA AGATCTATAT GAATTACACT TCCTCCGTAG
3001  GAGGAAGCAC AGGGGGAGAA TACCACTTCT CCCCCGGCGA CATAATGTAA ATGACGCAGT
3061  TTGCCTCGAA ATACTCCAGC TGCCCTGGAG TCATTTCCTT CATCCAATCT TCATCCGAGT
3121  TGGCGAGGAT TATTGTAGGC TTAGACTTCT TCTGCACCTT TTTCTTCTTA CCATACTTGG
3181  GGTTTACAAT GAAATCCCTC TGACAGCCAA CTAACTGTTT CCAACAAGGA CAGAATTTAA
3241  ACGGAATATC ATCTACGATG TTGTAGATTG CGTCTTCGTT GTATGAAGAC CAATCAACAT
3301  TATTTTGCCA GTAATTATGA ACCCCTAGGC TTCTGGCCCA AGTAGATTTT CCGGTTCTTG
3361  TTGGGCCGAC GATGTAGAGG CTCTGCTTTC TTGATCTTTC ATCTGATGAC TGGATACAGA
3421  ATCCATCCAT TGGAGGTCAG AAATTGCATC CTCGAGGGTA TAACAGGTAG GTTGAAGGAG
3481  CATGTAAGCT TCGGGACTAA CCTGGAAGAT GTTAGGCTGG AGCCAATCGT TGATTGACTC
3541  ATTACAAAGT AAATCAGGTG AGGAGGGTGG ATGAGGATTG GTGAACTCTT CCTGAATCTC
3601  AGGAAAAAGC TTATTTGCAG AGTATTCAAA ATACTGCAAT TTTGTGGACC AATCAAAGGG
3661  GAGCTCTTTC TGGATCATGG AGAGGTACTC TTCTTTGGAG GTAGCGTGTG AAATAATGTC
3721  TCGCATTATT TCATCTTTAG AAGGCTTTTT TTCCTTTACC TCTGAATCAG ATTTTCCTAG
3781  GAAGGGGGAC TTCCTAGGAA TGAAAGTACC TCTCTCAAAC ACAGCCAGAG GTTCCTTGAG
3841  AATGTAATCC CTCACTCTGT TAACTGACTT GGCACTCTGA ATATTTGGGT GAAACCCATT
3901  TATATCAAAG AACCTTGAGT CAGATATCCT TATCGGCTTC TCTGGCTGAA GCAATGCATG
3961  TAAATGCAAA CTTCCATCTT TATGTGCCTC TCGGGCACAT AGAATATATT TGGGAATCCA
4021  ACGAACGACG AGCTCCCAGA TCATCTGACA GGCGATTTCA GGATTTTCTG GACACTTTGG
4081  ATAGGTTAGG AACGTGTTAG CGTTCCTGTG TGAGAACTGA CGGTTGGATG AGGAGGAGGC
4141  CATAGCCGAC GACGGAGGTT GAGGCTGAGG GATGGCAGAC TGGGAGCTCC AAACTCTATA
4201  GTATACCCGT GCGCCTTCGA AATCCGCCGC TCCATTGTCT TATAGTGGTT GTAAATGGGC
4261  CGGACCGGGC CGGCCCAGCA GGAAAAGAAG GCGCGCACTA ATATTACCGC GCCTTCTTTT
4321  CCTGCGAGGG CCCGGGGTAG GGACCGAGCG CTTTGATTTA AAGCCTGGTT CTGCTTTGTA
4381  TGATTTATCT AAAGCAGCCC AATCTAAAGA AACCGGTCCC GGGCACTATA AATTGCCTAA
4441  CAAGTGCGAT TCATTCATGG ATCCTTTAAA CTCGAGTCTA GAGGGCCCAA TTCGCCCTAT
4501  AGTGAGTCGT ATTACAATTC ACTGGCCGTC GTTTTACAAC GTCGTGACTG GGAAAACCCT
4561  GGCGTTACCC AACTTAATCG CCTTGCAGCA CATCCCCCTT TCGCCAGCTG GCGTAATAGC
4621  GAAGAGGCCC GCACCGATCG CCCTTCCCAA CAGTTGCGCA GCCTATACGT ACGGCAGTTT
4681  AAGGTTTACA CCTATAAAAG AGAGAGCCGT TATCGTCTGT TTGTGGATGT ACAGAGTGAT
4741  ATTATTGACA CGCCGGGGCG ACGGATGGTG ATCCCCCTGG CCAGTGCACG TCTGCTGTCA
4801  GATAAAGTCT CCCGTGAACT TTACCCGGTG GTGCATATCG GGGATGAAAG CTGGCGCATG
4861  ATGACCACCG ATATGGCCAG TGTGCCGGTC TCCGTTATCG GGGAAGAAGT GGCTGATCTC
4921  AGCCACCGCG AAAATGACAT CAAAAACGCC ATTAACCTGA TGTTCTGGGG AATATAAATG
4981  TCAGGCCTGA ATGGCGAATG GACGCGCCCT GTAGCGGCGC ATTAAGCGCG CGGGTGTGGT
5041  GGTTACGCGC AGCGTGACCG CTACACTTGC CAGCGCCCTA GCGCCCGCTC CTTTCGCTTT
5101  CTTCCCTTCC TTTCTCGCCA CGTTCGCCGG CTTTCCCCGT CAAGCTCTAA ATCGGGGGCT
5161  CCCTTTAGGG TTCCGATTTA GAGCTTTACG GCACCTCGAC CGCAAAAAAC TTGATTTGGG
5221  TGATGGTTCA CGTAGTGGGC CATCGCCCTG ATAGACGGTT TTTCGCCCTT TGACGTTGGA
5281  GTCCACGTTC TTTAATAGTG GACTCTTGTT CCAAACTGGA ACAACACTCA ACCCTATCGC
5341  GGTCTATTCT TTTGATTTAT AAGGGATGTT GCCGATTTCG GCCTATTGGT TAAAAAATGA
5401  GCTGATTTAA CAAAAATTTT AACAAAATTC AGAAGAACTC GTCAAGAAGG CGATAGAAGG
5461  CGATGCGCTG CGAATCGGGA GCGGCGATAC CGTAAAGCAC GAGGAAGCGG TCAGCCCATT
5521  CGCCGCCAAG CTCTTCAGCA ATATCACGGG TAGCCAACGC TATGTCCTGA TAGCGGTCCG
5581  CCACACCCAG CCGGCCACAG TCGATGAATC CAGAAAAGCG GCCATTTTCC ACCATGATAT
5641  TCGGCAAGCA GGCATCGCCA TGGGTCACGA CGAGATCCTC GCCGTCGGGC ATGCTCGCCT
5701  TGAGCCTGGC GAACAGTTCG GCTGGCGCGA GCCCCTGATG CTCTTCGTCC AGATCATCCT
5761  GATCGACAAG ACCGGCTTCC ATCCGAGTAC GTGCTCGCTC GATGCGATGT TTCGCTTGGT
5821  GGTCGAATGG GCAGGTAGCC GGATCAAGCG TATGCAGCCG CCGCATTGCA TCAGCCATGA
5881  TGGATACTTT CTCGGCAGGA GCAAGGTGAG ATGACAGGAG ATCCTGCCCC GGCACTTCGC
5941  CCAATAGCAG CCAGTCCCTT CCCGCTTCAG TGACAACGTC GAGCACAGCT GCGCAAGGAA
6001  CGCCCGTCGT GGCCAGCCAC GATAGCCGCG CTGCCTCGTC TTGCAGTTCA TTCAGGGCAC
6061  CGGACAGGTC GGTCTTGACA AAAAGAACCG GGCGCCCCTG CGCTGACAGC CGGAACACGG
6121  CGGCATCAGA GCAGCCGATT GTCTGTTGTG CCCAGTCATA GCCGAATAGC CTCTCCACCC
6181  AAGCGGCCGG AGAACCTGCG TGCAATCCAT CTTGTTCAAT CATGCGAAAC GATCCTCATC
6241  CTGTCTCTTG ATCAGATCTT GATCCCCTGC GCCATCAGAT CCTTGGCGGC GAGAAAGCCA
```

Figure 9 (cont'd)

```
6301  TCCAGTTTAC TTTGCAGGGC TTCCCAACCT TACCAGAGGG CGCCCCAGCT GGCAATTCCG
6361  GTTCGCTTGC TGTCCATAAA ACCGCCCAGT CTAGCTATCG CCATGTAAGC CCACTGCAAG
6421  CTACCTGCTT TCTCTTTGCG CTTGCGTTTT CCCTTGTCCA GATAGCCCAG TAGCTGACAT
6481  TCATCCGGGG TCAGCACCGT TTCTGCGGAC TGGCTTTCTA CGTGAAAAGG ATCTAGGTGA
6541  AGATCCTTTT TGATAATCTC ATGACCAAAA TCCCTTAACG TGAGTTTTCG TTCCACTGAG
6601  CGTCAGACCC CGTAGAAAAG ATCAAAGGAT CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA
6661  TCTGCTGCTT GCAAACAAAA AAACCACCGC TACCAGCGGT GGTTTGTTTG CCGGATCAAG
6721  AGCTACCAAC TCTTTTTCCG AAGGTAACTG GCTTCAGCAG AGCGCAGATA CCAAATACTG
6781  TCCTTCTAGT GTAGCCGTAG TTAGGCCACC ACTTCAAGAA CTCTGTAGCA CCGCCTACAT
6841  ACCTCGCTCT GCTAATCCTG TTACCAGTGG CTGCTGCCAG TGGCGATAAG TCGTGTCTTA
6901  CCGGGTTGGA CTCAAGACGA TAGTTACCGG ATAAGGCGCA GCGGTCGGGC TGAACGGGGG
6961  GTTCGTGCAC ACAGCCCAGC TTGGAGCGAA CGACCTACAC CGAACTGAGA TACCTACAGC
7021  GTGAGCTATG AGAAAGCGCC ACGCTTCCCG AAGGGAGAAA GGCGGACAGG TATCCGGTAA
7081  GCGGCAGGGT CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGGAAAC GCCTGGTATC
7141  TTTATAGTCC TGTCGGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTTG TGATGCTCGT
7201  CAGGGGGGCG GAGCCTATGG AAAAACGCCA GCAACGCGGC CTTTTTACGG TTCCTGGGCT
7261  TTTGCTGGCC TTTTGCTCAC ATGTTCTTTC CTGCGTTATC CCCTGATTCT GTGGATAACC
7321  GTATTACCGC CTTTGAGTGA GCTGATACCG CTCGCCGCAG CCGAACGACC GAGCGCAGCG
7381  AGTCAGTGAG CGAGGAAGCG GAAG
```

SEQUENCE LISTING

<110> LARGE SCALE BIOLOGY CORPORATION

<120> COMPOSITIONS AND METHODS FOR INHIBITING GENE EXPRESSION

<130> 008010177PC00

<140> To Be Assigned
<141> 2001-04-04

<150> 09/545,574
<151> 2000-04-07

<160> 14

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 27
<212> DNA
<213> Cauliflower mosaic virus

<400> 1
tttgaattcg tcaacatggt ggagcac 27

<210> 2
<211> 31
<212> DNA
<213> Cauliflower mosaic virus

<400> 2
tttgtcgacg tcctctccaa atgaaatgaa c 31

<210> 3
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> zeocin resistance gene

<400> 3
cccgtcgact taattaagcg gccgcgttta caatttcgcc tgatgc 46

<210> 4
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> zeocin resistance gene

<400> 4
cccctcgagt taattaagcg gccgcctcaa aaaggatctt cacctag 47

<210> 5
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> nopaline synthase gene (nos) terminator sequence

<400> 5
tttctcgagc gaatttcccc gatcgttcaa ac 32

<210> 6
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> nopauline synthase (nos) terminator sequence

<400> 6
tttactagtc ccgatctagt aacatagatg ac 32

<210> 7
<211> 29
<212> DNA
<213> maize

<400> 7
tttttaatta aggtccgcct gaattctcg 29

<210> 8
<211> 30
<212> DNA
<213> maize

<400> 8
tttttaatta acggcaaggc tcacagtttg 30

<210> 9
<211> 4881
<212> DNA
<213> Viral

<400> 9
agcgcccaat acgcaaaccg cctctccccg cgcgttggcc gattcattaa tgcagctggc 60
acgacaggtt tcccgactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc 120
tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa 180
ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac gccaagctat 240
ttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaccgag ctcggatcca 300
ctagtaacgg ccgccagtgt gctggaattc atgggcagac ccgtctgtac tttaagagtg 360
ttggcaacca gtaatgaata aaaactcccg ttttattata tttgatgaat gctgaaagct 420
tacattaata tgtcgtgcga tggcacgaaa aaacacacgc aaacaataca ggggggtagt 480
cggcgggcgg ctaagggtgg tgctcggcgg gcagaacatc gaaaaatcaa gatctatatg 540
aattacactt cctccgtagg aggaagcaca gggggagaat accacttctc ccccggcgac 600
ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct gccctggagt catttccttc 660
atccaatctt catccgagtt ggcgaggatt attgtaggct tagacttctt ctgcaccttt 720

```
ttcttcttac catacttggg gtttacaatg aaatccctct gacagccaac taactgtttc     780
caacaaggac agaatttaaa cggaatatca tctacgatgt tgtagattgc gtcttcgttg     840
tatgaagacc aatcaacatt attttgccag taattatgaa cccctaggct tctggcccaa     900
gtagattttc cggttcttgt tgggccgacg atgtagaggc tctgctttct tgatctttca     960
tctgatgact ggatacagaa tccatccatt ggaggtcaga aattgcatcc tcgagggtat    1020
aacaggtagg ttgaaggagc atgtaagctt cgggactaac ctggaagatg ttaggctgga    1080
gccaatcgtt gattgactca ttacaaagta aatcaggtga ggagggtgga tgaggattgg    1140
tgaactcttc ctgaatctca ggaaaaagct tatttgcaga gtattcaaaa tactgcaatt    1200
ttgtggacca atcaaagggg agctctttct ggatcatgga gaggtactct tctttggagg    1260
tagcgtgtga aataatgtct cgcattattt catctttaga aggctttttt tcctttacct    1320
ctgaatcaga ttttcctagg aaggggggact tcctaggaat gaaagtacct ctctcaaaca    1380
cagccagagg ttccttgaga atgtaatccc tcactctgtt aactgacttg gcactctgaa    1440
tatttgggtg aaacccattt atatcaaaga accttgagtc agatatcctt atcggcttct    1500
ctggctgaag caatgcatgt aaatgcaaac ttccatcttt atgtgcctct cgggcacata    1560
gaatatattt gggaatccaa cgaacgacga gctcccagat catctgacag gcgatttcag    1620
gattttctgg acactttgga taggttagga acgtgttagc gttcctgtgt gagaactgac    1680
ggttggatga ggaggaggcc atagccgacg acggaggttg aggctgaggg atggcagact    1740
gggagctcca aactctatag tatacccgtg cgccttcgaa atccgccgct ccattgtctt    1800
atagtggttg taaatgggcc ggaccgggcc ggcccagcag gaaaagaagg cgcgcactaa    1860
tattaccgcg ccttcttttc ctgcgagggc ccggtaggga ccgagcgctt tgatttaaag    1920
cctggttctg ctttgcggcc gctcgagcat gcatctagag ggcccaattc gccctatagt    1980
gagtcgtatt acaattcact ggccgtcgtt ttacaacgtc gtgactggga aaaccctggc    2040
gttacccaac ttaatcgcct tgcagcacat ccccctttcg ccagctggcg taatagcgaa    2100
gaggcccgca ccgatcgccc ttcccaacag ttgcgcagcc tatacgtacg gcagtttaag    2160
gtttacacct ataaaagaga gagccgttat cgtctgtttg tggatgtaca gagtgatatt    2220
attgacacgc cggggcgacg gatggtgatc cccctggcca gtgcacgtct gctgtcagat    2280
aaagtctccc gtgaacttta cccggtggtg catatcgggg atgaaagctg gcgcatgatg    2340
accaccgata tggccagtgt gccggtctcc gttatcgggg aagaagtggc tgatctcagc    2400
caccgcgaaa atgacatcaa aaacgccatt aacctgatgt tctggggaat ataaatgtca    2460
ggcctgaatg gcgaatggac gcgccctgta gcggcgcatt aagcgcgcgg gtgtggtggt    2520
tacgcgcagc gtgaccgcta cacttgccag cgccctagcg cccgctcctt tcgctttctt    2580
cccttccttt ctcgccacgt tcgccggctt tccccgtcaa gctctaaatc gggggctccc    2640
tttagggttc cgatttagag ctttacggca cctcgaccgc aaaaaacttg atttgggtga    2700
tggttcacgt agtgggccat cgccctgata gacggttttt cgccctttga cgttggagtc    2760
cacgttcttt aatagtggac tcttgttcca aactggaaca acactcaacc ctatcgcggt    2820
ctattctttt gatttataag ggatgttgcc gatttcggcc tattggttaa aaaatgagct    2880
gatttaacaa aaattttaac aaaattcaga agaactcgtc aagaaggcga tagaaggcga    2940
tgcgctgcga atcgggagcg gcgataccgt aaagcacgag gaagcggtca gcccattcgc    3000
cgccaagctc ttcagcaata tcacgggtag ccaacgctat gtcctgatag cggtccgcca    3060
cacccagccg gccacagtcg atgaatccag aaaagcggcc attttccacc atgatattcg    3120
gcaagcaggc atcgccatgg gtcacgacga gatcctcgcc gtcgggcatg ctcgccttga    3180
gcctggcgaa cagttcggct ggcgcgagcc cctgatgctc ttcgtccaga tcatcctgat    3240
cgacaagacc ggcttccatc cgagtacgtg ctcgctcgat gcgatgtttc gcttggtggt    3300
cgaatgggca ggtagccgga tcaagcgtat gcagccgccg cattgcatca gccatgatgg    3360
atactttctc ggcaggagca aggtgagatg acaggagatc ctgccccggc acttcgccca    3420
atagcagcca gtcccttccc gcttcagtga caacgtcgag cacagctgcg caaggaacgc    3480
ccgtcgtggc cagccacgat agccgcgctg cctcgtcttg cagttcattc agggcaccgg    3540
acaggtcggt cttgacaaaa agaaccgggc gcccctgcgc tgacagccgg aacacggcgg    3600
catcagagca gccgattgtc tgttgtgccc agtcatagcc gaatagcctc tccacccaag    3660
cggccggaga acctgcgtgc aatccatctt gttcaatcat gcgaaacgat cctcatcctg    3720
tctcttgatc agatcttgat cccctgcgcc atcagatcct tggcggcgag aaagccatcc    3780
agtttacttt gcagggcttc ccaaccttac cagagggcgc cccagctggc aattccggtt    3840
cgcttgctgt ccataaaacc gcccagtcta gctatcgcca tgtaagccca ctgcaagcta    3900
cctgctttct ctttgcgctt gcgttttccc ttgtccagat agcccagtag ctgacattca    3960
tccggggtca gcaccgtttc tgcggactgg ctttctacgt gaaaaggatc taggtgaaga    4020
tccttttga taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt    4080
cagaccccgt agaaaagatc aaaggatctt cttgagatcc ttttttttctg cgcgtaatct    4140
```

gctgcttgca aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc 4200
taccaactct ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc 4260
ttctagtgta gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc 4320
tcgctctgct aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg 4380
ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggggtt 4440
cgtgcacaca gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg 4500
agctatgaga aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg 4560
gcagggtcgg aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggtatcttt 4620
atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag 4680
gggggcggag cctatggaaa aacgccagca acgcggcctt tttacggttc ctgggctttt 4740
gctggccttt tgctcacatg ttctttcctg cgttatcccc tgattctgtg gataaccgta 4800
ttaccgcctt tgagtgagct gataccgctc gccgcagccg aacgaccgag cgcagcgagt 4860
cagtgagcga ggaagcggaa g 4881

<210> 10
<211> 3413
<212> DNA
<213> viral

<400> 10
agcgcccaat acgcaaaccg cctctccccg cgcgttggcc gattcattaa tgcagctggc 60
acgacaggtt tcccgactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc 120
tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa 180
ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac gccaagctat 240
ttaggtgaca ctatagaata ctcaagctat gcatcaagct tgggcccggt agggaccgag 300
cgctttgatt taaagcctgg ttctgctttg tatgatttat ctaaagcagc ccaatctaaa 360
gaaaccggtc ccgggcacta taaattgcct aacaagtgcg attcattcat ggatccttta 420
aactcgagtc tagagggccc gaattctgca gatatccatc acactggcgg ccgctcgagc 480
atgcatctag agggcccaat tcgccctata gtgagtcgta ttacaattca ctggccgtcg 540
ttttacaacg tcgtgactgg gaaaaccctg gcgttaccca acttaatcgc cttgcagcac 600
atccccctttt cgccagctgg cgtaatagcg aagaggcccg caccgatcgc ccttcccaac 660
agttgcgcag cctatacgta cggcagttta aggtttacac ctataaaaga gagagccgtt 720
atcgtctgtt tgtggatgta cagagtgata ttattgacac gccggggcga cggatggtga 780
tccccctggc cagtgcacgt ctgctgtcag ataaagtctc ccgtgaactt tacccggtgg 840
tgcatatcgg ggatgaaagc tggcgcatga tgaccaccga tatggccagt gtgccggtct 900
ccgttatcgg ggaagaagtg gctgatctca gccaccgcga aaatgacatc aaaaacgcca 960
ttaacctgat gttctgggga atataaatgt caggcctgaa tggcgaatgg acgcgccctg 1020
tagcggcgca ttaagcgcgc gggtgtggtg gttacgcgca gcgtgaccgc tacacttgcc 1080
agcgccctag cgcccgctcc tttcgctttc ttcccttcct ttctcgccac gttcgccggc 1140
tttccccgtc aagctctaaa tcgggggctc cctttagggt tccgatttag agctttacgg 1200
cacctcgacc gcaaaaaact tgatttgggt gatggttcac gtagtgggcc atcgccctga 1260
tagacggttt ttcgcccttt gacgttggag tccacgttct ttaatagtgg actcttgttc 1320
caaactggaa caacactcaa ccctatcgcg gtctattctt ttgatttata agggatgttg 1380
ccgatttcgg cctattggtt aaaaaatgag ctgatttaac aaaaatttta acaaaattca 1440
gaagaactcg tcaagaaggc gatagaaggc gatgcgctgc gaatcgggag cggcgatacc 1500
gtaaagcacg aggaagcggt cagcccattc gccgccaagc tcttcagcaa tatcacgggt 1560
agccaacgct atgtcctgat agcggtccgc cacacccagc cggccacagt cgatgaatcc 1620
agaaaagcgg ccattttcca ccatgatatt cggcaagcag gcatcgccat gggtcacgac 1680
gagatcctcg ccgtcgggca tgctcgcctt gagcctggcg aacagttcgg ctggcgcgag 1740
cccctgatgc tcttcgtcca gatcatcctg atcgacaaga ccggcttcca tccgagtacg 1800
tgctcgctcg atgcgatgtt tcgcttggtg gtcgaatggg caggtagccg gatcaagcgt 1860
atgcagccgc cgcattgcat cagccatgat ggatactttc tcggcaggag caaggtgaga 1920
tgacaggaga tcctgccccg gcacttcgcc caatagcagc cagtcccttc ccgcttcagt 1980
gacaacgtcg agcacagctg cgcaaggaac gcccgtcgtg gccagccacg atagccgcgc 2040
tgcctcgtct tgcagttcat tcagggcacc ggacaggtcg gtcttgacaa aaagaaccgg 2100
gcgcccctgc gctgacagcc ggaacacggc ggcatcagag cagccgattg tctgttgtgc 2160
ccagtcatag ccgaatagcc tctccaccca agcggccgga gaacctgcgt gcaatccatc 2220

```
ttgttcaatc atgcgaaacg atcctcatcc tgtctcttga tcagatcttg atcccctgcg    2280
ccatcagatc cttggcggcg agaaagccat ccagtttact ttgcagggct tcccaacctt    2340
accagagggc gccccagctg gcaattccgg ttcgcttgct gtccataaaa ccgcccagtc    2400
tagctatcgc catgtaagcc cactgcaagc tacctgcttt ctctttgcgc ttgcgtttcc    2460
ccttgtccag atagcccagt agctgacatt catccggggt cagcaccgtt tctgcggact    2520
ggctttctac gtgaaaagga tctaggtgaa gatccttttt gataatctca tgaccaaaat    2580
cccttaacgt gagttttcgt tccactgagc gtcagacccc gtagaaaaga tcaaaggatc    2640
ttcttgagat cctttttttc tgcgcgtaat ctgctgcttg caaacaaaaa aaccaccgct    2700
accagcggtg gtttgtttgc cggatcaaga gctaccaact ctttttccga aggtaactgg    2760
cttcagcaga gcgcagatac caaatactgt ccttctagtg tagccgtagt taggccacca    2820
cttcaagaac tctgtagcac cgcctacata cctcgctctg ctaatcctgt taccagtggc    2880
tgctgccagt ggcgataagt cgtgtcttac cgggttggac tcaagacgat agttaccgga    2940
taaggcgcag cggtcgggct gaacgggggg ttcgtgcaca cagcccagct tggagcgaac    3000
gacctacacc gaactgagat acctacagcg tgagctatga gaaagcgcca cgcttcccga    3060
agggagaaag gcggacaggt atccggtaag cggcagggtc ggaacaggag agcgcacgag    3120
ggagcttcca gggggaaacg cctggtatct ttatagtcct gtcgggtttc gccacctctg    3180
acttgagcgt cgatttttgt gatgctcgtc agggggggcgg agcctatgga aaaacgccag    3240
caacgcggcc tttttacggt tcctgggctt ttgctggcct tttgctcaca tgttctttcc    3300
tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagtgag ctgataccgc    3360
tcgccgcagc cgaacgaccg agcgcagcga gtcagtgagc gaggaagcgg aag            3413
```

<210> 11
<211> 4961
<212> DNA
<213> Viral

<400> 11

```
agcgcccaat acgcaaaccg cctctccccg cgcgttggcc gattcattaa tgcagctggc      60
acgacaggtt tcccgactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc     120
tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa     180
ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac gccaagctat     240
ttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaccgag ctcggatcca     300
ctagtaacgg ccgccagtgt gctggaattc atgggcagac ccgtctgtac tttaagagtg     360
ttggcaacca gtaatgaata aaaactcccg ttttattata tttgatgaat gctgaaagct     420
tacattaata tgtcgtgcga tggcacgaaa aaacacacgc aaacaataca ggggggtagt     480
cggcgggcgg ctaagggtgg tgctcggcgg gcagaacatc gaaaaatcaa gatctatatg     540
aattacactt cctccgtagg aggaagcaca gggggagaat accacttctc ccccggcgac     600
ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct gccctggagt catttccttc     660
atccaatctt catccgagtt ggcgaggatt attgtaggct tagacttctt ctgcaccttt     720
ttcttcttac catacttggg gtttacaatg aaatccctct gacagccaac taactgtttc     780
caacaaggac agaatttaaa cggaatatca tctacgatgt tgtagattgc gtcttcgttg     840
tatgaagacc aatcaacatt attttgccag taattatgaa cccctaggct tctggcccaa     900
gtagattttc cggttcttgt tgggccgacg atgtagaggc tctgctttct tgatctttca     960
tctgatgact ggatacagaa tccatccatt ggaggtcaga aattgcatcc tcgagggtat    1020
aacaggtagg ttgaaggagc atgtaagctt cgggactaac ctggaagatg ttaggctgga    1080
gccaatcgtt gattgactca ttacaaagta aatcaggtga ggagggtgga tgaggattgg    1140
tgaactcttc ctgaatctca ggaaaaagct tatttgcaga gtattcaaaa tactgcaatt    1200
ttgtggacca atcaaagggg agctctttct ggatcatgga gaggtactct tctttggagg    1260
tagcgtgtga aataatgtct cgcattattt catctttaga aggctttttt tcctttacct    1320
ctgaatcaga ttttcctagg aaggggggact tcctaggaat gaaagtacct ctctcaaaca    1380
cagccagagg ttccttgaga atgtaatccc tcactctgtt aactgacttg gcactctgaa    1440
tatttgggtg aaacccattt atatcaaaga accttgagtc agatatcctt atcggcttct    1500
ctggctgaag caatgcatgt aaatgcaaac ttccatcttt atgtgcctct cgggcacata    1560
gaatatattt gggaatccaa cgaacgacga gctcccagat catctgacag gcgatttcag    1620
gattttctgg acactttgga taggttagga acgtgttagc gttcctgtgt gagaactgac    1680
ggttggatga ggaggaggcc atagccgacg acggaggttg aggctgaggg atggcagact    1740
gggagctcca aactctatag tatacccgtg cgccttcgaa atccgccgct ccattgtctt    1800
```

```
atagtggttg taaatgggcc ggaccgggcc ggcccagcag gaaaagaagg cgcgcactaa    1860
tattaccgcg ccttcttttc ctgcgagggc ccggtaggga ccgagcgctt tgatttaaag    1920
cctggttctg ctttgtatga tttatctaaa gcagcccaat ctaaagaaac cggtcccggg    1980
cactataaat tgcctaacaa gtgcgattca ttcatggatc ctttaaactc gagtctagag    2040
ggcccaattc gccctatagt gagtcgtatt acaattcact ggccgtcgtt ttacaacgtc    2100
gtgactggga aaaccctggc gttacccaac ttaatcgcct tgcagcacat ccccctttcg    2160
ccagctggcg taatagcgaa gaggcccgca ccgatcgccc ttcccaacag ttgcgcagcc    2220
tatacgtacg gcagtttaag gtttacacct ataaaagaga gagccgttat cgtctgtttg    2280
tggatgtaca gagtgatatt attgacacgc cggggcgacg gatggtgatc cccctggcca    2340
gtgcacgtct gctgtcagat aaagtctccc gtgaacttta cccggtggtg catatcgggg    2400
atgaaagctg gcgcatgatg accaccgata tggccagtgt gccggtctcc gttatcgggg    2460
aagaagtggc tgatctcagc caccgcgaaa atgacatcaa aaacgccatt aacctgatgt    2520
tctggggaat ataaatgtca ggcctgaatg gcgaatggac gcgccctgta gcggcgcatt    2580
aagcgcgcgg gtgtggtggt tacgcgcagc gtgaccgcta cacttgccag cgccctagcg    2640
cccgctcctt tcgctttctt cccttccttt ctcgccacgt tcgccggctt tccccgtcaa    2700
gctctaaatc gggggctccc tttagggttc cgatttagag ctttacggca cctcgaccgc    2760
aaaaaacttg atttgggtga tggttcacgt agtgggccat cgccctgata gacggttttt    2820
cgcccctttga cgttggagtc cacgttcttt aatagtggac tcttgttcca aactggaaca    2880
acactcaacc ctatcgcggt ctattctttt gatttataag ggatgttgcc gatttcggcc    2940
tattggttaa aaaatgagct gatttaacaa aaattttaac aaaattcaga agaactcgtc    3000
aagaaggcga tagaaggcga tgcgctgcga atcgggagcg gcgataccgt aaagcacgag    3060
gaagcggtca gcccattcgc cgccaagctc ttcagcaata tcacgggtag ccaacgctat    3120
gtcctgatag cggtccgcca cacccagccg gccacagtcg atgaatccag aaaagcggcc    3180
attttccacc atgatattcg gcaagcaggc atcgccatgg gtcacgacga gatcctcgcc    3240
gtcgggcatg ctcgccttga gcctggcgaa cagttcggct ggcgcgagcc cctgatgctc    3300
ttcgtccaga tcatcctgat cgacaagacc ggcttccatc cgagtacgtg ctcgctcgat    3360
gcgatgtttc gcttggtggt cgaatgggca ggtagccgga tcaagcgtat gcagccgccg    3420
cattgcatca gccatgatgg atactttctc ggcaggagca aggtgagatg acaggagatc    3480
ctgccccggc acttcgccca atagcagcca gtcccttccc gcttcagtga caacgtcgag    3540
cacagctgcg caaggaacgc ccgtcgtggc cagccacgat agccgcgctg cctcgtcttg    3600
cagttcattc agggcaccgg acaggtcggt cttgacaaaa agaaccgggc gcccctgcgc    3660
tgacagccgg aacacggcgg catcagagca gccgattgtc tgttgtgccc agtcatagcc    3720
gaatagcctc tccacccaag cggccggaga acctgcgtgc aatccatctt gttcaatcat    3780
gcgaaacgat cctcatcctg tctcttgatc agatcttgat cccctgcgcc atcagatcct    3840
tggcggcgag aaagccatcc agtttacttt gcagggcttc ccaaccttac cagagggcgc    3900
cccagctggc aattccggtt cgcttgctgt ccataaaacc gcccagtcta gctatcgcca    3960
tgtaagccca ctgcaagcta cctgctttct ctttgcgctt gcgttttccc ttgtccagat    4020
agcccagtag ctgacattca tccggggtca gcaccgtttc tgcggactgg ctttctacgt    4080
gaaaaggatc taggtgaaga tccttttttga taatctcatg accaaaatcc cttaacgtga    4140
gttttcgttc cactgagcgt cagaccccgt agaaaagatc aaaggatctt cttgagatcc    4200
ttttttttctg cgcgtaatct gctgcttgca aacaaaaaaa ccaccgctac cagcggtggt    4260
ttgtttgccg gatcaagagc taccaactct ttttccgaag gtaactggct tcagcagagc    4320
gcagatacca aatactgtcc ttctagtgta gccgtagtta ggccaccact tcaagaactc    4380
tgtagcaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg ctgccagtgg    4440
cgataagtcg tgtcttaccg ggttggactc aagacgatag ttaccggata aggcgcagcg    4500
gtcgggctga acgggggggtt cgtgcacaca gcccagcttg gagcgaacga cctacaccga    4560
actgagatac ctacagcgtg agctatgaga aagcgccacg cttcccgaag ggagaaaggc    4620
ggacaggtat ccggtaagcg gcagggtcgg aacaggagag cgcacgaggg agcttccagg    4680
gggaaacgcc tggtatcttt atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg    4740
atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca acgcggcctt    4800
tttacggttc ctgggctttt gctggccttt tgctcacatg ttctttcctg cgttatcccc    4860
tgattctgtg gataaccgta ttaccgcctt tgagtgagct gataccgctc gccgcagccg    4920
aacgaccgag cgcagcgagt cagtgagcga ggaagcggaa g                        4961
```

<210> 12
<211> 6309
<212> DNA

<213> Viral

<400> 12

agcgcccaat acgcaaaccg cctctccccg cgcgttggcc gattcattaa tgcagctggc 60
acgacaggtt tcccgactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc 120
tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa 180
ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac gccaagctat 240
ttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaccgag ctcggatcca 300
ctagtcccga tctagtaaca tagatgacac cgcgcgcgat aatttatcct agtttgcgcg 360
ctatattttg ttttctatcg cgtattaaat gtataattgc gggactctaa tcataaaaac 420
ccatctcata aataacgtca tgcattacat gttaattatt acatgcttaa cgtaattcaa 480
cagaaattat atgataatca tcgacagacc ggcaacagga ttcaatctta agaaacttta 540
ttgccaaatg tttgaacgat cggggaaatt cgctcgagtt aattaagcgg ccgcctcaaa 600
aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttagcacgt gtcagtcctg 660
ctcctcggcc acgaagtgca cgcagttgcc ggccgggtcg cgcagggcga actcccgccc 720
ccacggctgc tcgccgatct cggtcatggc cggcccggag gcgtcccgga agttcgtgga 780
cacgacctcc gaccactcgg cgtacagctc gtccaggccg cgcacccaca cccaggccag 840
ggtgttgtcc ggcaccacct ggtcctggac cgcgctgatg aacagggtca cgtcgtcccg 900
gaccacaccg gcgaagtcgt cctccacgaa gtcccgggag aacccgagcc ggtcggtcca 960
gaactcgacc gctccggcga cgtcgcgcgc ggtgagcacc ggaacggcac tggtcaactt 1020
ggccatggtg gccctcctca cgtgctatta ttgaagcatt tatcagggtt attgtctcat 1080
gagcggatac atatttgaat gtatttagaa aaataaacaa atagggggttc cgcgcacatt 1140
tccccgaaaa gtgccacctg tatgcggtgt gaaataccgc acagatgcgt aaggagaaaa 1200
taccgcatca ggcgaaattg taaacgcggc cgcttaatta agtcgacgtc ctctccaaat 1260
gaaatgaact tccttatata gaggaagggt cttgcgaagg atagtgggat tgtgcgtcat 1320
cccttacgtc agtggagata tcacatcaat ccacttgctt tgaagacgtg gttggaacgt 1380
cttcttttc cacgtagctc ctcgtgggtg ggggtccatc tttgggacca ctgtcggcag 1440
aggcatcttg aacgatagcc tttccttatc gcaatgatgg catttgtagg tgccaccttc 1500
cttttctact gtccttttga tgaagtgaca gatagctggg caatggaatc cgaggaggtt 1560
tcccgatatt acccttttgtt gaaaagtctc aatagccctt tggtcttctg agactgtatc 1620
tttgatattc ttggagtaga cgagagagtg tcgtgctcca ccatgttgac gaattcatgg 1680
gcagacccgt ctgtacttta agagtgttgg caaccagtaa tgaataaaaa ctcccgtttt 1740
attatatttg atgaatgctg aaagcttaca ttaatatgtc gtgcgatggc acgaaaaaac 1800
acacgcaaac aatacagggg ggtagtcggc gggcggctaa gggtggtgct cggcgggcag 1860
aacatcgaaa aatcaagatc tatatgaatt acacttcctc cgtaggagga agcacagggg 1920
gagaatacca cttctccccc ggcgacataa tgtaaatgac gcagtttgcc tcgaaatact 1980
ccagctgccc tggagtcatt tccttcatcc aatcttcatc cgagttggcg aggattattg 2040
taggcttaga cttcttctgc acctttttct tcttaccata cttggggttt acaatgaaat 2100
ccctctgaca gccaactaac tgtttccaac aaggacagaa tttaaacgga atatcatcta 2160
cgatgttgta gattgcgtct tcgttgtatg aagaccaatc aacattattt tgccagtaat 2220
tatgaacccc taggcttctg gcccaagtag attttccggt tcttgttggg ccgacgatgt 2280
agaggctctg ctttcttgat ctttcatctg atgactggat acagaatcca tccattggag 2340
gtcagaaatt gcatcctcga gggtataaca ggtaggttga aggagcatgt aagcttcggg 2400
actaacctgg aagatgttag gctggagcca atcgttgatt gactcattac aaagtaaatc 2460
aggtgaggag ggtggatgag gattggtgaa ctcttcctga atctcaggaa aaagcttatt 2520
tgcagagtat tcaaaatact gcaattttgt ggaccaatca aaggggagct ctttctggat 2580
catggagagg tactcttctt tggaggtagc gtgtgaaata atgtctcgca ttatttcatc 2640
tttagaaggc ttttttcct ttacctctga atcagatttt cctaggaagg gggacttcct 2700
aggaatgaaa gtacctctct caaacacagc cagaggttcc ttgagaatgt aatccctcac 2760
tctgttaact gacttggcac tctgaatatt tgggtgaaac ccatttatat caaagaacct 2820
tgagtcagat atccttatcg gcttctctgg ctgaagcaat gcatgtaaat gcaaacttcc 2880
atctttatgt gcctctcggg cacatagaat atatttggga atccaacgaa cgacgagctc 2940
ccagatcatc tgacaggcga tttcaggatt ttctggacac tttggatagg ttaggaacgt 3000
gttagcgttc ctgtgtgaga actgacggtt ggatgaggag gaggccatag ccgacgacgg 3060
aggttgaggc tgagggatgg cagactggga gctccaaact ctatagtata cccgtgcgcc 3120
ttcgaaatcc gccgctccat tgtcttatag tggttgtaaa tgggccggac cgggccggcc 3180
cagcaggaaa agaaggcgcg cactaatatt accgcgcctt cttttcctgc gagggcccgg 3240

ggtagggacc gagcgctttg atttaaagcc tggttctgct ttgtatgatt tatctaaagc 3300
agcccaatct aaagaaaccg gtcccgggca ctataaattg cctaacaagt gcgattcatt 3360
catggatcct ttaaactcga gtctagaggg cccaattcgc cctatagtga gtcgtattac 3420
aattcactgg ccgtcgtttt acaacgtcgt gactgggaaa accctggcgt tacccaactt 3480
aatcgccttg cagcacatcc ccctttcgcc agctggcgta atagcgaaga ggcccgcacc 3540
gatcgccctt cccaacagtt gcgcagccta tacgtacggc agtttaaggt ttacacctat 3600
aaaagagaga gccgttatcg tctgtttgtg gatgtacaga gtgatattat tgacacgccg 3660
gggcgacgga tggtgatccc cctggccagt gcacgtctgc tgtcagataa agtctcccgt 3720
gaactttacc cggtggtgca tatcggggat gaaagctggc gcatgatgac caccgatatg 3780
gccagtgtgc cggtctccgt tatcggggaa gaagtggctg atctcagcca ccgcgaaaat 3840
gacatcaaaa acgccattaa cctgatgttc tggggaatat aaatgtcagg cctgaatggc 3900
gaatggacgc gccctgtagc ggcgcattaa gcgcgcgggt gtggtggtta cgcgcagcgt 3960
gaccgctaca cttgccagcg ccctagcgcc cgctcctttc gctttcttcc cttcctttct 4020
cgccacgttc gccggctttc cccgtcaagc tctaaatcgg gggctccctt tagggttccg 4080
atttagagct ttacggcacc tcgaccgcaa aaaacttgat ttgggtgatg gttcacgtag 4140
tgggccatcg ccctgataga cggtttttcg ccctttgacg ttggagtcca cgttctttaa 4200
tagtggactc ttgttccaaa ctggaacaac actcaaccct atcgcggtct attcttttga 4260
tttataaggg atgttgccga tttcggccta ttggttaaaa aatgagctga tttaacaaaa 4320
attttaacaa aattcagaag aactcgtcaa gaaggcgata gaaggcgatg cgctgcgaat 4380
cgggagcggc gataccgtaa agcacgagga agcggtcagc ccattcgccg ccaagctctt 4440
cagcaatatc acgggtagcc aacgctatgt cctgatagcg gtccgccaca cccagccggc 4500
cacagtcgat gaatccagaa aagcggccat tttccaccat gatattcggc aagcaggcat 4560
cgccatgggt cacgacgaga tcctcgccgt cgggcatgct cgccttgagc ctggcgaaca 4620
gttcggctgg cgcgagcccc tgatgctctt cgtccagatc atcctgatcg acaagaccgg 4680
cttccatccg agtacgtgct cgctcgatgc gatgtttcgc ttggtggtcg aatgggcagg 4740
tagccggatc aagcgtatgc agccgccgca ttgcatcagc catgatggat actttctcgg 4800
caggagcaag gtgagatgac aggagatcct gccccggcac ttcgcccaat agcagccagt 4860
cccttcccgc ttcagtgaca acgtcgagca cagctgcgca aggaacgccc gtcgtggcca 4920
gccacgatag ccgcgctgcc tcgtcttgca gttcattcag ggcaccggac aggtcggtct 4980
tgacaaaaag aaccgggcgc ccctgcgctg acagccggaa cacggcggca tcagagcagc 5040
cgattgtctg ttgtgcccag tcatagccga atagcctctc cacccaagcg gccggagaac 5100
ctgcgtgcaa tccatcttgt tcaatcatgc gaaacgatcc tcatcctgtc tcttgatcag 5160
atcttgatcc cctgcgccat cagatccttg gcggcgagaa agccatccag tttactttgc 5220
agggcttccc aaccttacca gagggcgccc cagctggcaa ttccggttcg cttgctgtcc 5280
ataaaaccgc ccagtctagc tatcgccatg taagcccact gcaagctacc tgctttctct 5340
ttgcgcttgc gttttccctt gtccagatag cccagtagct gacattcatc cggggtcagc 5400
accgtttctg cggactggct ttctacgtga aaaggatcta ggtgaagatc ctttttgata 5460
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag 5520
aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa 5580
caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt 5640
ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc 5700
cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa 5760
tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa 5820
gacgatagtt accggataag gcgcagcggt cgggctgaac gggggttcg tgcacacagc 5880
ccagcttgga gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa 5940
gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc agggtcggaa 6000
caggagagcg cacgagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg 6060
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc 6120
tatggaaaaa cgccagcaac gcggcctttt tacggttcct gggcttttgc tggccttttg 6180
ctcacatgtt ctttcctgcg ttatcccctg attctgtgga taaccgtatt accgcctttg 6240
agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg 6300
aagcggaag 6309

<210> 13
<211> 8043

<212> DNA
<213> Viral

<400> 13

```
agcgcccaat acgcaaaccg cctctccccg cgcgttggcc gattcattaa tgcagctggc      60
acgacaggtt tcccgactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc     120
tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa     180
ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac gccaagctat     240
ttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaccgag ctcggatcca     300
ctagtaacgg ccgccagtgt gctggaattc atgggcagac ccgtctgtac tttaagagtg     360
ttggcaacca gtaatgaata aaaactcccg ttttattata tttgatgaat gctgaaagct     420
tacattaata tgtcgtgcga tggcacgaaa aaacacacgc aaacaataca ggggggtagt     480
cggcgggcgg ctaagggtgg tgctcggcgg gcagaacatc gaaaaatcaa gatctatatg     540
aattacactt cctccgtagg aggaagcaca gggggagaat accacttctc ccccggcgac     600
ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct gccctggagt catttccttc     660
atccaatctt catccgagtt ggcgaggatt attgtaggct tagacttctt ctgcaccttt     720
ttcttcttac catacttggg gtttacaatg aaatccctct gacagccaac taactgtttc     780
caacaaggac agaatttaaa cggaatatca tctacgatgt tgtagattgc gtcttcgttg     840
tatgaagacc aatcaacatt attttgccag taattatgaa cccctaggct tctggcccaa     900
gtagattttc cggttcttgt tgggccgacg atgtagaggc tctgctttct tgatctttca     960
tctgatgact ggatacagaa tccatccatt ggaggtcaga aattgcatcc tcgagggtat    1020
aacaggtagg ttgaaggagc atgtaagctt cgggactaac ctggaagatg ttaggctgga    1080
gccaatcgtt gattgactca ttacaaagta aatcaggtga ggagggtgga tgaggattgg    1140
tgaactcttc ctgaatctca ggaaaaagct tatttgcaga gtattcaaaa tactgcaatt    1200
ttgtggacca atcaaagggg agctctttct ggatcatgga gaggtactct tctttggagg    1260
tagcgtgtga aataatgtct cgcattattt catctttaga aggcttttct tcctttacct    1320
ctgaatcaga ttttcctagg aaggggggact tcctaggaat gaaagtacct ctctcaaaca    1380
cagccagagg ttccttgaga atgtaatccc tcactctgtt aactgacttg gcactctgaa    1440
tatttgggtg aaacccattt atatcaaaga accttgagtc agatatcctt atcggcttct    1500
ctggctgaag caatgcatgt aaatgcaaac ttccatcttt atgtgcctct cgggcacata    1560
gaatatattt gggaatccaa cgaacgacga gctcccagat catctgacag gcgatttcag    1620
gattttctgg acactttgga taggttagga acgtgttagc gttcctgtgt gagaactgac    1680
ggttggatga ggaggaggcc atagccgacg acggaggttg aggctgaggg atggcagact    1740
gggagctcca aactctatag tatacccgtg cgccttcgaa atccgccgct ccattgtctt    1800
atagtggttg taaatgggcc ggaccgggcc ggcccagcag gaaaagaagg cgcgcactaa    1860
tattaccgcg ccttcttttc ctgcgagggc ccggtaggga ccgagcgctt tgatttaaag    1920
cctggttctg ctttgtatga tttatctaaa gcagcccaat ctaaagaaac cggtcccggg    1980
cactataaat tgcctaacaa gtgcgattca ttcatggatc ctttaaactc gagtctagtc    2040
ccgatctagt aacatagatg acaccgcgcg cgataattta tcctagtttg cgcgctatat    2100
tttgttttct atcgcgtatt aaatgtataa ttgcgggact ctaatcataa aaacccatct    2160
cataaataac gtcatgcatt acatgttaat tattacatgc ttaacgtaat tcaacagaaa    2220
ttatatgata atcatcgaca gaccggcaac aggattcaat cttaagaaac tttattgcca    2280
aatgtttgaa cgatcgggga aattcgctcg agttaattaa gcggccgcct caaaaaggat    2340
cttcacctag atccttttaa attaaaaatg aagttttagc acgtgtcagt cctgctcctc    2400
ggccacgaag tgcacgcagt tgccggccgg gtcgcgcagg gcgaactccc gcccccacgg    2460
ctgctcgccg atctcggtca tggccggccc ggaggcgtcc cggaagttcg tggacacgac    2520
ctccgaccac tcggcgtaca gctcgtccag gccgcgcacc cacacccagg ccagggtgtt    2580
gtccggcacc acctggtcct ggaccgcgct gatgaacagg gtcacgtcgt cccggaccac    2640
accggcgaag tcgtcctcca cgaagtcccg ggagaacccg agccggtcgg tccagaactc    2700
gaccgctccg gcgacgtcgc gcgcggtgag caccggaacg gcactggtca acttggccat    2760
ggtggccctc ctcacgtgct attattgaag catttatcag ggttattgtc tcatgagcgg    2820
atacatattt gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg    2880
aaaagtgcca cctgtatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc    2940
atcaggcgaa attgtaaacg cggccgctta attaagtcga cgtcctctcc aaatgaaatg    3000
aacttcctta tatagaggaa gggtcttgcg aaggatagtg ggattgtgcg tcatccctta    3060
cgtcagtgga gatatcacat caatccactt gctttgaaga cgtggttgga acgtcttctt    3120
tttccacgta gctcctcgtg ggtgggggtc catctttggg accactgtcg gcagaggcat    3180
```

```
cttgaacgat agcctttcct tatcgcaatg atggcatttg taggtgccac cttccttttc    3240
tactgtcctt ttgatgaagt gacagatagc tgggcaatgg aatccgagga ggtttcccga    3300
tattaccctt tgttgaaaag tctcaatagc cctttggtct tctgagactg tatctttgat    3360
attcttggag tagacgagag agtgtcgtgc tccaccatgt tgacgaattc atgggcagac    3420
ccgtctgtac tttaagagtg ttggcaacca gtaatgaata aaaactcccg ttttattata    3480
tttgatgaat gctgaaagct tacattaata tgtcgtgcga tggcacgaaa aaacacacgc    3540
aaacaataca ggggggtagt cggcgggcgg ctaagggtgg tgctcggcgg gcagaacatc    3600
gaaaaatcaa gatctatatg aattacactt cctccgtagg aggaagcaca gggggagaat    3660
accacttctc ccccggcgac ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct    3720
gccctggagt catttccttc atccaatctt catccgagtt ggcgaggatt attgtaggct    3780
tagacttctt ctgcaccttt ttcttcttac catacttggg gtttacaatg aaatccctct    3840
gacagccaac taactgtttc caacaaggac agaatttaaa cggaatatca tctacgatgt    3900
tgtagattgc gtcttcgttg tatgaagacc aatcaacatt attttgccag taattatgaa    3960
cccctaggct tctggcccaa gtagattttc cggttcttgt tgggccgacg atgtagaggc    4020
tctgctttct tgatctttca tctgatgact ggatacagaa tccatccatt ggaggtcaga    4080
aattgcatcc tcgagggtat aacaggtagg ttgaaggagc atgtaagctt cgggactaac    4140
ctggaagatg ttaggctgga gccaatcgtt gattgactca ttacaaagta aatcaggtga    4200
ggagggtgga tgaggattgg tgaactcttc ctgaatctca ggaaaaagct tatttgcaga    4260
gtattcaaaa tactgcaatt ttgtggacca atcaaagggg agctctttct ggatcatgga    4320
gaggtactct tctttggagg tagcgtgtga aataatgtct cgcattattt catctttaga    4380
aggctttttt tcctttacct ctgaatcaga ttttcctagg aaggggggact tcctaggaat    4440
gaaagtacct ctctcaaaca cagccagagg ttccttgaga atgtaatccc tcactctgtt    4500
aactgacttg gcactctgaa tatttgggtg aaacccattt atatcaaaga accttgagtc    4560
agatatcctt atcggcttct ctggctgaag caatgcatgt aaatgcaaac ttccatcttt    4620
atgtgcctct cgggcacata gaatatattt gggaatccaa cgaacgacga gctcccagat    4680
catctgacag gcgatttcag gattttctgg acactttgga taggttagga acgtgttagc    4740
gttcctgtgt gagaactgac ggttggatga ggaggaggcc atagccgacg acggaggttg    4800
aggctgaggg atggcagact gggagctcca aactctatag tatacccgtg cgccttcgaa    4860
atccgccgct ccattgtctt atagtggttg taaatgggcc ggaccgggcc ggcccagcag    4920
gaaaagaagg cgcgcactaa tattaccgcg ccttcttttc ctgcgagggc ccggggtagg    4980
gaccgagcgc tttgatttaa agcctggttc tgctttgtat gatttatcta aagcagccca    5040
atctaaagaa accggtcccg ggcactataa attgcctaac aagtgcgatt cattcatgga    5100
tcctttaaac tcgagtctag agggcccaat tcgccctata gtgagtcgta ttacaattca    5160
ctggccgtcg ttttacaacg tcgtgactgg gaaaaccctg gcgttaccca acttaatcgc    5220
cttgcagcac atccccctttt cgccagctgg cgtaatagcg aagaggcccg caccgatcgc    5280
ccttcccaac agttgcgcag cctatacgta cggcagttta aggtttacac ctataaaaga    5340
gagagccgtt atcgtctgtt tgtggatgta cagagtgata ttattgacac gccggggcga    5400
cggatggtga tccccctggc cagtgcacgt ctgctgtcag ataaagtctc ccgtgaactt    5460
tacccggtgg tgcatatcgg ggatgaaagc tggcgcatga tgaccaccga tatggccagt    5520
gtgccggtct ccgttatcgg ggaagaagtg gctgatctca gccaccgcga aaatgacatc    5580
aaaaacgcca ttaacctgat gttctgggga atataaatgt caggcctgaa tggcgaatgg    5640
acgcgccctg tagcggcgca ttaagcgcgc gggtgtggtg gttacgcgca gcgtgaccgc    5700
tacacttgcc agcgccctag cgcccgctcc tttcgctttc ttcccttcct ttctcgccac    5760
gttcgccggc tttccccgtc aagctctaaa tcgggggctc cctttagggt tccgatttag    5820
agctttacgg cacctcgacc gcaaaaaact tgatttgggt gatggttcac gtagtgggcc    5880
atcgccctga tagacggttt ttcgcccttt gacgttggag tccacgttct ttaatagtgg    5940
actcttgttc caaactggaa caacactcaa ccctatcgcg gtctattctt ttgatttata    6000
agggatgttg ccgatttcgg cctattggtt aaaaaatgag ctgatttaac aaaaatttta    6060
acaaaattca gaagaactcg tcaagaaggc gatagaaggc gatgcgctgc gaatcgggag    6120
cggcgatacc gtaaagcacg aggaagcggt cagcccattc gccgccaagc tcttcagcaa    6180
tatcacgggt agccaacgct atgtcctgat agcggtccgc cacacccagc cggccacagt    6240
cgatgaatcc agaaaagcgg ccattttcca ccatgatatt cggcaagcag gcatcgccat    6300
gggtcacgac gagatcctcg ccgtcgggca tgctcgcctt gagcctggcg aacagttcgg    6360
ctggcgcgag cccctgatgc tcttcgtcca gatcatcctg atcgacaaga ccggcttcca    6420
tccgagtacg tgctcgctcg atgcgatgtt tcgcttggtg gtcgaatggg caggtagccg    6480
gatcaagcgt atgcagccgc cgcattgcat cagccatgat ggatactttc tcggcaggag    6540
caaggtgaga tgacaggaga tcctgccccg gcacttcgcc caatagcagc cagtcccttc    6600
```

```
ccgcttcagt gacaacgtcg agcacagctg cgcaaggaac gcccgtcgtg gccagccacg    6660
atagccgcgc tgcctcgtct tgcagttcat tcagggcacc ggacaggtcg gtcttgacaa    6720
aaagaaccgg gcgcccctgc gctgacagcc ggaacacggc ggcatcagag cagccgattg    6780
tctgttgtgc ccagtcatag ccgaatagcc tctccaccca agcggccgga gaacctgcgt    6840
gcaatccatc ttgttcaatc atgcgaaacg atcctcatcc tgtctcttga tcagatcttg    6900
atcccctgcg ccatcagatc cttggcggcg agaaagccat ccagtttact ttgcagggct    6960
tcccaacctt accagagggc gccccagctg gcaattccgg ttcgcttgct gtccataaaa    7020
ccgcccagtc tagctatcgc catgtaagcc cactgcaagc tacctgcttt ctctttgcgc    7080
ttgcgttttc ccttgtccag atagcccagt agctgacatt catccggggt cagcaccgtt    7140
tctgcggact ggctttctac gtgaaaagga tctaggtgaa gatccttttt gataatctca    7200
tgaccaaaat cccttaacgt gagttttcgt tccactgagc gtcagacccc gtagaaaaga    7260
tcaaaggatc ttcttgagat cctttttttc tgcgcgtaat ctgctgcttg caaacaaaaa    7320
aaccaccgct accagcggtg gtttgtttgc cggatcaaga gctaccaact ctttttccga    7380
aggtaactgg cttcagcaga gcgcagatac caaatactgt ccttctagtg tagccgtagt    7440
taggccacca cttcaagaac tctgtagcac cgcctacata cctcgctctg ctaatcctgt    7500
taccagtggc tgctgccagt ggcgataagt cgtgtcttac cgggttggac tcaagacgat    7560
agttaccgga taaggcgcag cggtcgggct gaacgggggg ttcgtgcaca cagcccagct    7620
tggagcgaac gacctacacc gaactgagat acctacagcg tgagctatga gaaagcgcca    7680
cgcttcccga agggagaaag gcggacaggt atccggtaag cggcagggtc ggaacaggag    7740
agcgcacgag ggagcttcca gggggaaacg cctggtatct ttatagtcct gtcgggtttc    7800
gccacctctg acttgagcgt cgatttttgt gatgctcgtc agggggggcgg agcctatgga    7860
aaaacgccag caacgcggcc ttttacggt tcctgggctt ttgctggcct tttgctcaca    7920
tgttctttcc tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagtgag    7980
ctgataccgc tcgccgcagc cgaacgaccg agcgcagcga gtcagtgagc gaggaagcgg    8040
aag                                                                 8043
```

<210> 14
<211> 7404
<212> DNA
<213> Viral

<400> 14

```
agcgcccaat acgcaaaccg cctctccccg cgcgttggcc gattcattaa tgcagctggc      60
acgacaggtt tcccgactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc     120
tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa     180
ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac gccaagctat     240
ttaggtgaca ctatagaata ctcaagctat gcatcaagct tggtaccgag ctcggatcca     300
ctagtaacgg ccgccagtgt gctggaattc atgggcagac ccgtctgtac tttaagagtg     360
ttggcaacca gtaatgaata aaaactcccg ttttattata tttgatgaat gctgaaagct     420
tacattaata tgtcgtgcga tggcacgaaa aaacacacgc aaacaataca gggggggtagt    480
cggcgggcgg ctaagggtgg tgctcggcgg gcagaacatc gaaaaatcaa gatctatatg     540
aattacactt cctccgtagg aggaagcaca gggggagaat accacttctc ccccggcgac     600
ataatgtaaa tgacgcagtt tgcctcgaaa tactccagct gccctggagt catttccttc     660
atccaatctt catccgagtt ggcgaggatt attgtaggct tagacttctt ctgcaccttt     720
ttcttcttac catacttggg gtttacaatg aaatccctct gacagccaac taactgtttc     780
caacaaggac agaatttaaa cggaatatca tctacgatgt tgtagattgc gtcttcgttg     840
tatgaagacc aatcaacatt attttgccag taattatgaa cccctaggct tctggcccaa     900
gtagattttc cggttcttgt tgggccgacg atgtagaggc tctgctttct tgatctttca     960
tctgatgact ggatacagaa tccatccatt ggaggtcaga aattgcatcc tcgagggtat    1020
aacaggtagg ttgaaggagc atgtaagctt cgggactaac ctggaagatg ttaggctgga    1080
gccaatcgtt gattgactca ttacaaagta aatcaggtga ggagggtgga tgaggattgg    1140
tgaactcttc ctgaatctca ggaaaaagct tatttgcaga gtattcaaaa tactgcaatt    1200
ttgtggacca atcaaagggg agctctttct ggatcatgga gaggtactct tctttggagg    1260
tagcgtgtga aataatgtct cgcattattt catctttaga aggctttttt tcctttacct    1320
ctgaatcaga ttttcctagg aaggggggact tcctaggaat gaaagtacct ctctcaaaca    1380
```

```
cagccagagg ttccttgaga atgtaatccc tcactctgtt aactgacttg gcactctgaa    1440
tatttgggtg aaacccattt atatcaaaga accttgagtc agatatcctt atcggcttct    1500
ctggctgaag caatgcatgt aaatgcaaac ttccatcttt atgtgcctct cgggcacata    1560
gaatatattt gggaatccaa cgaacgacga gctcccagat catctgacag gcgatttcag    1620
gattttctgg acactttgga taggttagga acgtgttagc gttcctgtgt gagaactgac    1680
ggttggatga ggaggaggcc atagccgacg acggaggttg aggctgaggg atggcagact    1740
gggagctcca aactctatag tatacccgtg cgccttcgaa atccgccgct ccattgtctt    1800
atagtggttg taaatgggcc ggaccgggcc ggcccagcag gaaaagaagg cgcgcactaa    1860
tattaccgcg ccttcttttc ctgcgagggc ccggtaggga ccgagcgctt tgatttaaag    1920
cctggttctg ctttgtatga tttatctaaa gcagcccaat ctaaagaaac cggtcccggg    1980
cactataaat tgcctaacaa gtgcgattca ttcatggatc ctttaaactc gagtctagtc    2040
ccgatctagt aacatagatg acaccgcgcg cgataattta tcctagtttg cgcgctatat    2100
tttgttttct atcgcgtatt aaatgtataa ttgcgggact ctaatcataa aaacccatct    2160
cataaataac gtcatgcatt acatgttaat tattacatgc ttaacgtaat tcaacagaaa    2220
ttatatgata atcatcgaca gaccggcaac aggattcaat cttaagaaac tttattgcca    2280
aatgtttgaa cgatcgggga aattcgctcg agttaattaa gcggccgctt aattaagtcg    2340
acgtcctctc caaatgaaat gaacttcctt atatagagga agggtcttgc gaaggatagt    2400
gggattgtgc gtcatccctt acgtcagtgg agatatcaca tcaatccact tgctttgaag    2460
acgtggttgg aacgtcttct ttttccacgt agctcctcgt gggtgggggt ccatctttgg    2520
gaccactgtc ggcagaggca tcttgaacga tagcctttcc ttatcgcaat gatggcattt    2580
gtaggtgcca ccttcctttt ctactgtcct tttgatgaag tgacagatag ctgggcaatg    2640
gaatccgagg aggtttcccg atattaccct ttgttgaaaa gtctcaatag ccctttggtc    2700
ttctgagact gtatctttga tattcttgga gtagacgaga gagtgtcgtg ctccaccatg    2760
ttgacgaatt catgggcaga cccgtctgta ctttaagagt gttggcaacc agtaatgaat    2820
aaaaactccc gttttattat atttgatgaa tgctgaaagc ttacattaat atgtcgtgcg    2880
atggcacgaa aaaacacacg caaacaatac agggggggtag tcggcgggcg gctaagggtg    2940
gtgctcggcg ggcagaacat cgaaaaatca agatctatat gaattacact tcctccgtag    3000
gaggaagcac agggggagaa taccacttct cccccggcga cataatgtaa atgacgcagt    3060
ttgcctcgaa atactccagc tgccctggag tcatttcctt catccaatct tcatccgagt    3120
tggcgaggat tattgtaggc ttagacttct tctgcacctt tttcttctta ccatacttgg    3180
ggtttacaat gaaatccctc tgacagccaa ctaactgttt ccaacaagga cagaatttaa    3240
acggaatatc atctacgatg ttgtagattg cgtcttcgtt gtatgaagac caatcaacat    3300
tattttgcca gtaattatga acccctaggc ttctggccca agtagatttt ccggttcttg    3360
ttgggccgac gatgtagagg ctctgctttc ttgatctttc atctgatgac tggatacaga    3420
atccatccat tggaggtcag aaattgcatc ctcgagggta taacaggtag gttgaaggag    3480
catgtaagct tcgggactaa cctggaagat gttaggctgg agccaatcgt tgattgactc    3540
attacaaagt aaatcaggtg aggagggtgg atgaggattg gtgaactctt cctgaatctc    3600
aggaaaaagc ttatttgcag agtattcaaa atactgcaat tttgtggacc aatcaaaggg    3660
gagctctttc tggatcatgg agaggtactc ttctttggag gtagcgtgtg aaataatgtc    3720
tcgcattatt tcatctttag aaggcttttt ttcctttacc tctgaatcag attttcctag    3780
gaaggggggac ttcctaggaa tgaaagtacc tctctcaaac acagccagag gttccttgag    3840
aatgtaatcc ctcactctgt taactgactt ggcactctga atatttgggt gaaacccatt    3900
tatatcaaag aaccttgagt cagatatcct tatcggcttc tctggctgaa gcaatgcatg    3960
taaatgcaaa cttccatctt tatgtgcctc tcgggcacat agaatatatt tgggaatcca    4020
acgaacgacg agctcccaga tcatctgaca ggcgatttca ggattttctg gacactttgg    4080
ataggttagg aacgtgttag cgttcctgtg tgagaactga cggttggatg aggaggaggc    4140
catagccgac gacggaggtt gaggctgagg gatggcagac tgggagctcc aaactctata    4200
gtatacccgt gcgccttcga aatccgccgc tccattgtct tatagtggtt gtaaatgggc    4260
cggaccgggc cggcccagca ggaaaagaag gcgcgcacta atattaccgc gccttctttt    4320
cctgcgaggg cccggggtag ggaccgagcg ctttgattta aagcctggtt ctgctttgta    4380
tgatttatct aaagcagccc aatctaaaga aaccggtccc gggcactata aattgcctaa    4440
caagtgcgat tcattcatgg atcctttaaa ctcgagtcta gagggcccaa ttcgccctat    4500
agtgagtcgt attacaattc actggccgtc gttttacaac gtcgtgactg ggaaaaccct    4560
ggcgttaccc aacttaatcg ccttgcagca catccccctt tcgccagctg gcgtaatagc    4620
gaagaggccc gcaccgatcg cccttcccaa cagttgcgca gcctatacgt acggcagttt    4680
aaggtttaca cctataaaag agagagccgt tatcgtctgt ttgtggatgt acagagtgat    4740
attattgaca cgccggggcg acggatggtg atcccccctgg ccagtgcacg tctgctgtca    4800
```

```
gataaagtct cccgtgaact ttacccggtg gtgcatatcg gggatgaaag ctggcgcatg    4860
atgaccaccg atatggccag tgtgccggtc tccgttatcg gggaagaagt ggctgatctc    4920
agccaccgcg aaaatgacat caaaaacgcc attaacctga tgttctgggg aatataaatg    4980
tcaggcctga atggcgaatg gacgcgccct gtagcggcgc attaagcgcg cgggtgtggt    5040
ggttacgcgc agcgtgaccg ctacacttgc cagcgcccta gcgcccgctc ctttcgcttt    5100
cttcccttcc tttctcgcca cgttcgccgg ctttccccgt caagctctaa atcgggggct    5160
ccctttaggg ttccgattta gagctttacg gcacctcgac cgcaaaaaac ttgatttggg    5220
tgatggttca cgtagtgggc atcgccctg atagacggtt tttcgccctt tgacgttgga    5280
gtccacgttc tttaatagtg gactcttgtt ccaaactgga acaacactca accctatcgc    5340
ggtctattct tttgatttat aagggatgtt gccgatttcg gcctattggt taaaaaatga    5400
gctgatttaa caaaaatttt aacaaaattc agaagaactc gtcaagaagg cgatagaagg    5460
cgatgcgctg cgaatcggga gcggcgatac cgtaaagcac gaggaagcgg tcagcccatt    5520
cgccgccaag ctcttcagca atatcacggg tagccaacgc tatgtcctga tagcggtccg    5580
ccacacccag ccggccacag tcgatgaatc cagaaaagcg gccattttcc accatgatat    5640
tcggcaagca ggcatcgcca tgggtcacga cgagatcctc gccgtcgggc atgctcgcct    5700
tgagcctggc gaacagttcg gctggcgcga gcccctgatg ctcttcgtcc agatcatcct    5760
gatcgacaag accggcttcc atccgagtac gtgctcgctc gatgcgatgt ttcgcttggt    5820
ggtcgaatgg gcaggtagcc ggatcaagcg tatgcagccg ccgcattgca tcagccatga    5880
tggatacttt ctcggcagga gcaaggtgag atgacaggag atcctgcccc ggcacttcgc    5940
ccaatagcag ccagtccctt cccgcttcag tgacaacgtc gagcacagct gcgcaaggaa    6000
cgcccgtcgt ggccagccac gatagccgcg ctgcctcgtc ttgcagttca ttcagggcac    6060
cggacaggtc ggtcttgaca aaaagaaccg ggcgcccctg cgctgacagc cggaacacgg    6120
cggcatcaga gcagccgatt gtctgttgtg cccagtcata gccgaatagc ctctccaccc    6180
aagcggccgg agaacctgcg tgcaatccat cttgttcaat catgcgaaac gatcctcatc    6240
ctgtctcttg atcagatctt gatcccctgc gccatcagat ccttggcggc gagaaagcca    6300
tccagtttac tttgcagggc ttcccaacct taccagaggg cgccccagct ggcaattccg    6360
gttcgcttgc tgtccataaa accgcccagt ctagctatcg ccatgtaagc ccactgcaag    6420
ctacctgctt tctctttgcg cttgcgtttt cccttgtcca gatagcccag tagctgacat    6480
tcatccgggg tcagcaccgt ttctgcggac tggctttcta cgtgaaaagg atctaggtga    6540
agatcctttt tgataatctc atgaccaaaa tcccttaacg tgagttttcg ttccactgag    6600
cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga tccttttttt ctgcgcgtaa    6660
tctgctgctt gcaaacaaaa aaaccaccgc taccagcggt ggtttgtttg ccggatcaag    6720
agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata ccaaatactg    6780
tccttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca ccgcctacat    6840
acctcgctct gctaatcctg ttaccagtgg ctgctgccag tggcgataag tcgtgtctta    6900
ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc tgaacggggg    6960
gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga tacctacagc    7020
gtgagctatg agaaagcgcc acgcttcccg aagggagaaa ggcggacagg tatccggtaa    7080
gcggcagggt cggaacagga gagcgcacga gggagcttcc agggggaaac gcctggtatc    7140
tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgatttttg tgatgctcgt    7200
caggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg ttcctgggct    7260
tttgctggcc ttttgctcac atgttctttc ctgcgttatc ccctgattct gtggataacc    7320
gtattaccgc ctttgagtga gctgataccg ctcgccgcag ccgaacgacc gagcgcagcg    7380
agtcagtgag cgaggaagcg gaag                                           7404
```